

UNIVERSITÉ DE MONTRÉAL

Synthèse de vues à partir d'images prises par des caméras
stéréoscopiques non calibrées

MOULAY ABDELLATIF AKHLOUFI

DÉPARTEMENT DE GÉNIE ÉLECTRIQUE ET DE GÉNIE INFORMATIQUE
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES (M.Sc.A.)
(GÉNIE ÉLECTRIQUE)

AVRIL 1999

© Moulay Abdellatif AKHLOUFI, 1999.



**National Library
of Canada**

**Acquisitions and
Bibliographic Services**

**395 Wellington Street
Ottawa ON K1A 0N4
Canada**

**Bibliothèque nationale
du Canada**

**Acquisitions et
services bibliographiques**

**395, rue Wellington
Ottawa ON K1A 0N4
Canada**

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-42893-1

Canada

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

Synthèse de vues à partir d'images prises par des caméras
stéréoscopiques non calibrées

présenté par: AKHLOUFI Moulay Abdellatif,

en vue de l'obtention du diplôme de: Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de:

HERVÉ Jean-Yves, Ph.D., président

COHEN Paul, Ph.D., membre et directeur de recherche

POLOTSKI Vladimir, Ph.D., membre et codirecteur de recherche

POULIN Pierre, Ph.D., membre

À ma famille et à ma chère do avec tout mon amour.

Remerciements

Je voudrais remercier mon directeur de recherche Paul Cohen, ainsi que mon codirecteur Vladimir Polotski pour leurs aide et conseils durant la réalisation de ce travail.

Je remercie également Messieurs Jean-Yves Hervé et Pierre Poulin qui m'ont fait l'honneur de participer au jury et d'avoir pris le temps d'évaluer mon travail.

Je tiens aussi à remercier mes collègues du Groupe de Recherche en Perception et Robotique.

Résumé

Ce travail présente une nouvelle approche pour la synthèse de nouvelles vues à partir d'images bidimensionnelles prises par des caméras stéréoscopiques non calibrées, et ce sans reconstruction tridimensionnelle explicite. Ainsi, seulement un petit nombre de points appariés sur deux images sources est utilisé pour générer la nouvelle vue. Aussi, pour obtenir un *rendu géométriquement valide* nous utilisons des outils de la géométrie projective qui permettent une description réaliste de la scène.

Dans un premier temps, nous présentons les notions de géométrie projective qui nous permettront de définir les relations entre les caméras dans la scène. Les notions de géométrie épipolaire sont mises à contribution pour introduire une nouvelle méthode de calcul de la *matrice fondamentale*, et ce pour décrire les propriétés inhérentes à une configuration d'une paire de caméras stéréoscopiques. Elle a comme caractéristiques d'être linéaire, ce qui permet de réduire le temps de calcul de façon appréciable, et d'exploiter les informations présentes dans la scène, à savoir les surfaces planes, pour obtenir de meilleurs résultats comparativement aux techniques classiques de calcul de la *matrice fondamentale*.

Dans un deuxième temps, les *invariants projectifs* sont étudiés et deux techniques pour les calculer à partir de l'appariement des points entre images sources sont présentées. Nous les utilisons ensuite pour décrire une représentation invariante de la scène en termes de birapports des points appariés dans deux images sources. Une étude comparative nous permet de montrer que le calcul de ces invariants donne des résultats très satisfaisants. Ici aussi nous exploitons la présence de surfaces planes dans la scène pour calculer de façon robuste une structure invariante de celle-ci.

Enfin, nous utilisons la géométrie épipolaire et les invariants projectifs pour synthétiser des nouvelles vues. Nous utilisons un petit nombre de points appariés pour obtenir une grille de points de contrôle sur chacune des images sources. Cette grille est alors reprojétée sur sa nouvelle position dans un troisième plan image en utilisant les techniques développées précédemment (matrice fondamentale et invariants projectifs). Le reste des points images est ensuite obtenu par interpolation des points de la grille à l'aide d'une technique de *texturage bidimensionnel*¹ en perspective. Ceci nous permet de faire un rendu basé sur images qui a comme caractéristique d'être rapide tout en préservant un bon réalisme pour les nouvelles vues générées. Notre approche est finalement illustrée sur plusieurs exemples de scènes réelles et synthétiques.

¹warping

Abstract

Without the computational cost of 3D reconstruction, this thesis presents an original approach for synthesis of a third view from two images captured by a non-calibrated stereo system. Only sparse correspondences between the source images are needed. The approach results in a realistic *geometrically-valid* rendering of the scene.

With the aid of projective geometry, geometrical relations characterizing a two cameras configuration are defined. A new technique for *fundamental matrix* computation is then presented. It exploits the available geometrical information about the scene, in particular, the planarity constraints in the epipolar equations. This technique is linear and computationally less expensive than classical methods.

Next, *projective invariants* are presented with two techniques for their computation. This yields an invariant representation of the scene in terms of cross-ratios of corresponding points in a pair of source images. Recognition of planar surfaces in the scene leads to a robust computation of the invariant representation.

The foregoing synthesis of a novel view depends on epipolar geometry and projective invariants. A small number of corresponding points is required to obtain a

grid of anchor points in each image plane. This grid is reprojected on its position in the third view using fundamental matrix and projective invariants. The remaining image points are rendered by perspective *image warping* to achieve a photo-realistic rendered view. The efficiency of the method is illustrated on images of synthetic and real scenes.

Table des matières

Dédicace	iv
Remerciements	v
Résumé	vi
Abstract	viii
Table des matières	x
Liste des tableaux	xiv
Liste des figures	xv
Liste des notations	xix
1 Introduction	1
2 Relations géométriques entre caméras	9
2.1 Géométrie projective en vision	10

2.1.1	Espaces projectifs	10
2.1.2	Le plan projectif	12
2.1.3	Coordonnées homogènes	13
2.1.4	La ligne projective	14
2.1.5	Transformations projectives	16
2.1.6	Transformation projective de lignes	18
2.1.7	Calcul de l'homographie	19
2.2	Modèle de la caméra	23
2.2.1	Modèle projectif	23
2.2.2	Modèle de projection perspective	24
2.3	Caméras stéréoscopiques	29
2.4	Calcul de la matrice fondamentale	35
2.4.1	Méthodes Linéaires	36
2.4.2	Méthodes non linéaires	46
2.5	Approche proposée : Exploitation de la contrainte de planarité	52
2.5.1	Description de la méthode proposée	55
2.5.2	Relation avec la matrice fondamentale	59
2.5.3	Cas d'un modèle de caméras stéréo calibrées	68
2.6	Conclusion	70
3	Invariants projectifs	72
3.1	Introduction aux invariants projectifs	73

3.1.1	Définition d'un invariant	73
3.1.2	Le birapport	74
3.1.3	Invariants projectifs des transformations de \mathcal{P}^3 dans \mathcal{P}^2	82
3.1.4	Coordonnées projectives	83
3.1.5	Dualité	88
3.1.6	Géométries affine et euclidienne	89
3.2	Invariant basé sur un point de référence et deux homographies	95
3.3	Invariant basé sur une ligne de référence et une seule homographie	100
3.4	Conclusion	105
4	Synthèse de vues	107
4.1	État de l'art en synthèse de vues	108
4.1.1	Synthèse par reconstruction tridimensionnelle explicite	108
4.1.2	Mosaïquage tridimensionnel	111
4.1.3	Rendu basé sur les images : Approches non géométriques	113
4.1.4	Rendu basé sur les images: Approches géométriques	115
4.1.5	Mosaïquage bidimensionnel	120
4.1.6	Approche utilisant une base de données d'images	122
4.2	Approche proposée	123
4.3	Techniques de transfert de primitives d'intérêt	124
4.3.1	Transfert à l'aide de la géométrie épipolaire	124
4.3.2	Transfert à l'aide d'un invariant projectif	129

4.4	Texturage bidimensionnel	132
4.4.1	Transformation affine	134
4.4.2	Transformation perspective	135
4.4.3	Transformation bilinéaire	136
4.4.4	Décomposition de l'image	137
4.5	Algorithme de synthèse de vues	140
4.6	Conclusion	140
5	Résultats expérimentaux	144
5.1	Géométrie épipolaire	144
5.1.1	Matrice fondamentale	145
5.1.2	Comparaisons	151
5.2	Invariants projectifs	158
5.3	Synthèse de vues	165
5.4	Conclusion	170
6	Conclusion	172
	Bibliographie	178

Liste des tableaux

5.1	Résultats de comparaison entre différentes techniques de calcul de la matrice fondamentale par la technique Fdiff	154
5.2	Facteur de qualité pour différentes techniques de calcul de la matrice fondamentale	156

Liste des figures

1.1	Description globale de l'approche proposée pour la synthèse d'une nouvelle vue	8
2.1	Modèle de plan projectif	13
2.2	Modèle d'obtention d'une ligne à partir de deux points dans le plan projectif	16
2.3	Colinéation entre deux plans projectifs	17
2.4	Modèle de projection perspective	25
2.5	Modèle de caméras stéréoscopiques	30
2.6	Géométrie épipolaire	35
2.7	Homographie entre surfaces planes	55
2.8	Contrainte de planarité et géométrie épipolaire	58
2.9	Contrainte définie par la matrice fondamentale et le point épipolaire .	61
3.1	L'ordre des points peut être changé par une transformation projective générale, il est préservé par une projection centrale	79

3.2	Birapport de quatre droites	81
3.3	Birapport d'un faisceau de quatre plans	82
3.4	Coordonnées projectives dans un plan	84
3.5	Birapport associé à cinq droites	90
3.6	Birapport en géométrie euclidienne	93
3.7	Transformations euclidiennes de l'espace et la conique absolue	95
3.8	Invariant basé sur un point et deux homographies	101
3.9	Birapport de deux points et deux lignes	102
3.10	Invariant basé sur une ligne de référence et une seule homographie	105
4.1	Synthèse de vues par interpolation	116
4.2	Intersection des lignes épipolaires près du plan trifocal	118
4.3	Contraintes de trilinearité	120
4.4	Mosaïquage bidimensionnel cylindrique	122
4.5	Géométrie de trois caméras dans l'espace	125
4.6	Transfert de points à l'aide de la géométrie épipolaire	126
4.7	Transfert de lignes à l'aide de la géométrie épipolaire	128
4.8	Transfert proche du plan trifocal: choix des points P et Q loin du plan trifocal	129
4.9	Transfert proche du plan trifocal à l'aide de la géométrie épipolaire	130
4.10	Transfert à l'aide d'un invariant projectif	131
4.11	Description de l'approche proposée	143

5.1	Résultats des différentes techniques d'estimation de la géométrie épipolaire sur la paire stéréo "Mire"	146
5.2	Résultats des différentes techniques d'estimation de la géométrie épipolaire sur la paire stéréo "Bâtiment"	147
5.3	Résultats des différentes techniques d'estimation de la géométrie épipolaire sur la paire stéréo "Bureau"	148
5.4	Résultats des différentes techniques d'estimation de la géométrie épipolaire sur la paire stéréo "Maison"	149
5.5	Résultats des différentes techniques d'estimation de la géométrie épipolaire sur la paire stéréo "Lab"	150
5.6	Différence entre matrices fondamentales	152
5.7	Paire stéréoscopique "Mire" utilisée pour le calcul des invariants projectifs	159
5.8	Paire stéréoscopique "Bureau" utilisée pour le calcul des invariants projectifs	160
5.9	Résultats de calcul des invariants projectifs pour la paire "Mire" . . .	161
5.10	Résultats de calcul des invariants projectifs pour la paire "Bureau" .	162
5.11	Images sources pour la paire d'images "Bâtiment"	165
5.12	Images sources pour la paire d'images "Maison"	166
5.13	Images sources pour la paire d'images "Lab"	166
5.14	Image synthétisée - "Bâtiment"	167

5.15 Images synthétisées:(a) les images réelles; (b) images synthétisées par géométrie épipolaire; (c) images synthétisées par invariants projectifs	
- "Maison"	168
5.16 Images synthétisées:(a) les images réelles; (b) images synthétisées par géométrie épipolaire; (c) images synthétisées par invariants projectifs	
- "Lab"	169

Liste des notations

- \mathcal{P}^n : espace projectif de dimension n ;
- $P = [x_1, \dots, x_{n+1}]^t$: un vecteur de $(n + 1)$ éléments représentant un point de l'espace projectif de dimension n , et x_i ses coordonnées homogènes;
- λ : scalaire non nul;
- $l = [l_1, l_2, l_3]^t$: ligne projective, et l_i ses coordonnées homogènes;
- $\overline{m} = [x_1, x_2, x_3]^t$: un point du plan projectif;
- H : matrice 3×3 de transformation homogène entre deux plans ou homographie;
- \min : fonction de minimisation;
- $P = [X, Y, Z]$: point dans l'espace 3D;
- $_{cam}$: indice représentant le repère de la caméra;
- $_w$: indice représentant le repère monde ou repère de la scène;
- f : la focale de la caméra;

- \otimes : le produit vectoriel;
- R : matrice de rotation;
- t : vecteur de translation;
- $\Omega(t)$: matrice antisymétrique du vecteur t ;
- M : point de l'espace projectif 3D;
- $.l$ ou $.l'$: representent le plan image de gauche;
- $.r$ ou $.r'$: representent le plan image de droite;
- E : matrice essentielle;
- F : matrice fondamentale;
- f_{ij} : éléments de la matrice fondamentale F ;
- ep : appelé point épipolaire;
- Π : plan dans l'espace 3D;
- $d(m_1, m_2)$: distance algébrique entre deux points m_1 et m_2 ;
- $\bar{k} = [A, B, C, D] = \frac{\overline{AC} \times \overline{BD}}{\overline{AD} \times \overline{BC}}$: birapport de quatre points;
- \overline{AB} : mesure algébrique de AB ou la distance entre les points A et B ;
- $j(\bar{k}) = \frac{(\bar{k}^2 - \bar{k} + 1)^3}{\bar{k}^2(\bar{k} - 1)}$: le J invariant;

- $L_{ox\Pi_1}$: projection de la ligne L_{ox} sur le plan Π_1 ;
- $l_{\pi_1}^r$: image de $L_{ox\Pi_1}$ dans le plan image de droite;
- $l_{\pi_1}^l$: image de $L_{ox\Pi_1}$ dans le plan image de gauche;
- $(l_1 x_1)$: la distance entre la ligne l_1 et le point x_1 ;
- ${}^{\Pi}L_{ref}$: projection de la ligne L_{ref} sur le plan Π_{ref} ;
- $e_{i,i+1}$: le point épipolaire sur le plan image i correspondant au centre optique de la caméra $i + 1$;
- $Fm = l$: la ligne epipolaire correspondant au point m .

Chapitre 1

Introduction

Depuis le début de la vision par ordinateur, les chercheurs se sont intéressés à l'analyse et l'interprétation de la géométrie de la scène, à partir des images prises par des caméras. L'objectif étant de représenter l'espace tridimensionnel par des entités géométriques d'une manière efficace. Une représentation de ce type devrait tenir compte de l'aspect dynamique de l'environnement, pour permettre d'exploiter cette information pour l'exécution de diverses tâches (par un robot équipé de caméras par exemple). Il est donc important de pouvoir produire une interprétation aussi complète que possible de la scène à partir seulement de l'information visuelle disponible. Un des domaines en vision qui vise à produire une représentation complète de la scène est la reconstruction de la structure tridimensionnelle d'une scène à partir seulement des images de celle-ci, et ce pour des applications diverses comme la planification, l'évitement d'obstacles, et plus récemment la réalité virtuelle entre autres. C'est

ce dernier domaine qui connaît un grand engouement ces dernières années, qui a poussé les deux communautés de vision et graphisme par ordinateur (infographie) à joindre leurs efforts pour trouver des solutions pratiques aux problèmes posés en réalité virtuelle. Des chercheurs de ces deux domaines traditionnellement séparés ont commencé récemment à travailler sur le développement d'outils permettant de produire un rendu réaliste à partir d'un ensemble d'images d'une même scène prises par des caméras réelles. Ce domaine s'appelle *rendu basé sur images* (McMillan et Bishop, 1995; Kang, 1997; Dyer, 1997; Lengyel, 1998). Il consiste en des techniques de synthèse de nouvelles vues à partir de vues connues et ce sans modèle tridimensionnel de la scène.

Dans les approches traditionnelles en vision par ordinateur la synthèse de vues était produite à partir d'un modèle tridimensionnel proprement reconstruit à partir de l'appariement établi entre les images de la scène. Une fois le modèle obtenu, on lui applique une texture extraite des images sources pour ajouter du réalisme. On peut ensuite placer une source de lumière pour obtenir un rendu amélioré. Après cette reconstruction, la scène peut être manipulée et on peut choisir la position d'une caméra virtuelle qui va capturer la scène du point de vue désiré. Cette caméra pourrait aussi bien être couplée à la tête d'un opérateur pour une visualisation interactive de différents points de vues de la scène reconstruite. Cette approche est encore utilisée de nos jours dans différentes applications (Berger, Simon, Petitjean et Wrobel-Dautcourt, 1996; Chang et Zakhor, 1995; Chang et Zakhor, 1997; Debevec et Malik, 1996). En

infographie une approche similaire est utilisée. D'abord un modèle de la scène est reconstruit à partir de plans de celle-ci ou à partir de données tridimensionnelles fournies par un senseur tridimensionnel. Puis l'ajout de la texture et des sources de lumière permet un rendu plus réaliste de la scène à partir d'un point de vue choisi en plaçant une caméra virtuelle dans la scène (Foley, Van Dam, Feiner et Hughes, 1990; Watt, 1993). Toutefois, ces deux approches ont quelques inconvénients :

1. La reconstruction tridimensionnelle en vision est très coûteuse en temps de calcul et est très sensible aux erreurs d'appariements.
2. La modélisation en infographie a un coût prohibitif en ressources homme-machine et dépend de la précision des résultats fournis par les senseurs tridimensionnels utilisés et/ou les plans disponibles de la scène.

Pour remédier à ces problèmes, un nouveau domaine a été introduit récemment, c'est le rendu basé sur images. L'objectif est de permettre la synthèse de nouvelles vues à partir seulement des images connues et ce sans reconstruction tridimensionnelle explicite. Ceci permet d'éviter le temps de calcul souvent prohibitif que nécessite la modélisation tridimensionnelle de la scène. Imaginons qu'à partir de quelques images prises sur un site historique nous puissions nous promener dans un environnement virtuel représentant ce site et ce sans avoir à faire un modèle tridimensionnel de celui-ci. Nous pourrions avoir différents points de vue en utilisant la synthèse de nouvelles vues à partir seulement de vues bidimensionnelles connues. Nous pourrions ainsi avoir accès à un tel environnement sur un ordinateur personnel sans avoir besoin

de matériel sophistiqué. Cette approche pourrait être intéressante pour la télérobotique par exemple, où un opérateur assis devant son écran d'ordinateur et qui veut manipuler un objet à distance dans une scène non modélisée a priori (Cohen, Hervé et Akhloufi, 1998), les caméras fournissant l'information visuelle prise sur le site de travail. L'opérateur pourrait en bougeant la tête avoir différentes vues par simple synthèse de nouvelles vues à partir des images fournies par les caméras sans avoir à déplacer ces dernières dans le site ou faire une reconstruction tridimensionnelle fastidieuse de la scène. Ainsi, les techniques de rendu basé sur images fournissent un outil pratique et intéressant dans ce sens, d'où l'intérêt croissant pour ce domaine ces dernières années.

Dans le travail que nous présentons dans ce mémoire, nous proposons de définir la scène non pas par un modèle tridimensionnel, mais par des vues bidimensionnelles réelles de celle-ci. En supprimant ainsi l'étape de modélisation et en accélérant l'étape de rendu nous pouvons produire de nouvelles vues dans un temps acceptable. En effet, les images disponibles contiennent les informations géométriques nécessaires, et les informations de texture sous forme déjà *rendue*.

L'information géométrique contenue dans les images nous permet dans le cadre de la géométrie projective de mieux représenter la scène en vue de la génération d'une nouvelle vue. Les outils de la géométrie projective permettent la synthèse d'une nouvelle vue satisfaisant les contraintes géométriques inhérentes à une configuration des caméras et de la scène. La vue ainsi obtenue sera plus proche de la vue qui serait

fournie par une caméra placée au même endroit que la caméra virtuelle.

Tout cela nous amène dans le cadre de ce travail à nous intéresser aux outils de géométrie projective qui permettent un rendu *géométriquement valide*, c'est-à-dire la géométrie épipolaire et les invariants projectifs. En ce sens, notre contribution se situera au niveau du développement de nouvelles méthodes basées sur l'exploitation des propriétés de la scène, plus particulièrement la présence de surfaces planes dans celle-ci. Cette contrainte de planarité est caractérisée par une relation de transformation homographique entre les projections sur les plans images d'une même surface plane. Cette transformation contient toute l'information nécessaire pour préserver la planarité d'une surface projetée.

Dans le cadre de ce travail nous contribuons à l'estimation de la géométrie épipolaire par une approche originale de calcul de la matrice fondamentale. La méthode proposée est linéaire et nécessite un temps de calcul faible comparativement aux méthodes existantes (Luong et Faugeras, 1994*b*; Deriche, Zhang, Luong et Faugeras, 1994; Luong et Faugeras, 1996; Zhang, 1996; Csurka, Zeller, Zhang et Faugeras, 1997; Faugeras, Luong et Maybank, 1992; Hartley, 1997*a*; Hartley, 1997*d*; Hartley, 1997*b*).

Au niveau des invariants projectifs, nous présenterons deux techniques de calcul de ces invariants à partir de deux images d'une paire stéréoscopique qui donnent des résultats très satisfaisants.

Finalement, pour la synthèse de nouvelles vues à partir d'images connues, nous

décrivons une approche originale qui, contrairement aux techniques de synthèse de vues existantes (Laveau et Faugeras, 1994; Seitz et Dyer, 1995*b*; Seitz et Dyer, 1996*a*; Avidan et Shashua, 1997; Kanade, 1996; Kanade, Narayanan et Rander, 1996), ne nécessite qu'un appariement épars entre les points d'intérêt présents dans les images sources pour générer une nouvelle vue de la scène. La méthode développée s'inspire des techniques d'infographie, plus particulièrement la Conception Assistée par Ordinateur (CAO) (Watt, 1993; Zhang, 1998*a*), où la scène est modélisée par un ensemble de plans ou surfaces de base qui sont associés ensemble pour définir la scène et sur lesquels on colle une texture pour ajouter plus de réalisme. Dans notre cas, les deux images stéréoscopiques sources sont décomposées en morceaux de surfaces planes à partir d'un petit nombre de points d'intérêt appariés, décrivant ainsi un maillage apparié sur les deux images. L'idée est de reprojeter les points appariés décrivant ce maillage, c'est-à-dire les points d'intérêt qui jouent le rôle de points de contrôle, dans une nouvelle vue par des techniques de géométrie projective que nous développerons dans le cadre de ce travail (matrice fondamentale et invariants projectifs). Une fois les points de contrôle placés sur le nouveau plan image, nous transférons ensuite le reste des points en utilisant une technique de *texturage bidimensionnel* en perspective qui permet de préserver la planarité des différents morceaux de surfaces composant les images après projection (figure 1.1). Nous évitons ainsi de faire un appariement dense entre les images de la paire stéréoscopique, qui peut être prohibitif en temps machine.

Organisation du travail

Le travail est organisé de la façon suivante: le chapitre 2 présente les relations géométriques fondamentales entre caméras. Nous donnons une brève introduction de la géométrie projective qui nous permet de définir les relations épipolaires entre caméras. Ensuite, une revue bibliographique des principales techniques de calcul de la matrice fondamentale est présentée, puis une nouvelle méthode est introduite. Elle repose sur l'existence de surfaces planes dans la scène pour faire le calcul de la matrice fondamentale. Le chapitre 3 donne un aperçu sur les invariants projectifs pour finir par une présentation de deux techniques pour calculer ces invariants à partir de l'appariement des primitives d'intérêt (points et lignes) sur une paire d'images stéréoscopiques. Au chapitre 4, les techniques de synthèse de vues sont introduites par une revue bibliographique qui décrit les principales méthodes existantes dans la littérature. Ensuite, une approche originale pour générer de nouvelles vues à partir seulement d'un appariement éparse entre les images est présentée. Le chapitre 5 présente les résultats expérimentaux illustrant la validité des méthodes proposées. Une brève conclusion résume ce travail. Nous y décrivons les contributions apportées, discutons des améliorations possibles et soulevons des questions qui peuvent être abordées dans de futurs travaux.

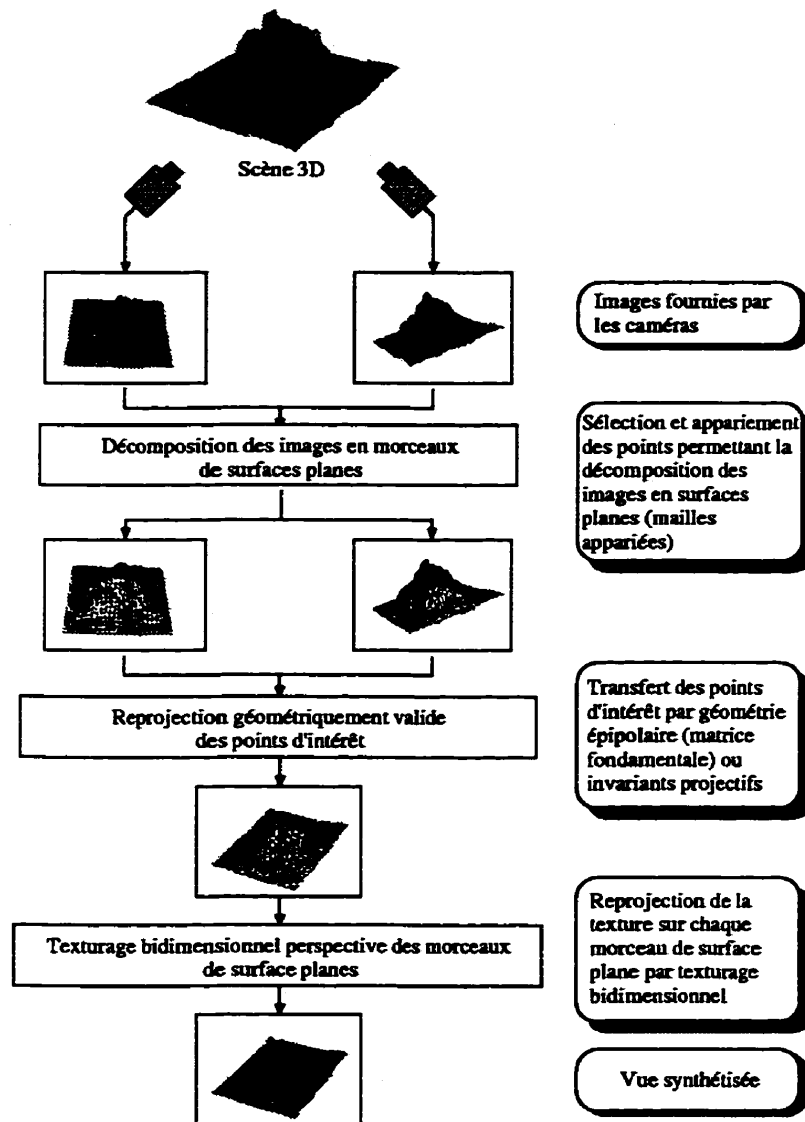


Figure 1.1 : Description globale de l'approche proposée pour la synthèse d'une nouvelle vue

Chapitre 2

Relations géométriques entre caméras

Dans ce chapitre nous décrivons les relations qui lient les coordonnées tridimensionnelles d'un point de la scène à celles, bidimensionnelles, de son image. La géométrie projective nous offre les outils pour représenter cette relation de manière élégante, permettant ainsi une représentation linéaire projective.

Le chapitre commence par une brève introduction à la géométrie projective (Semple et Kneebone, 1979), où nous allons décrire les résultats fondamentaux qui seront utiles par la suite.

Dans la seconde partie nous allons aborder l'aspect géométrique de la formation d'image. Nous décrivons les modèles de projection que nous utiliserons et l'apport de la géométrie projective dans ce sens.

Enfin, nous présenterons les notions de géométrie épipolaire et de matrice fondamentale. Nous commencerons par une revue des principales méthodes pour estimer la matrice fondamentale. Nous présenterons ensuite une propriété importante reliant l'homographie à la matrice fondamentale, et finirons par proposer une nouvelle approche basée sur l'exploitation de cette propriété et la connaissance que nous avons de la scène.

2.1 Géométrie projective en vision

2.1.1 Espaces projectifs

Un point appartenant à un espace projectif de dimension n , \mathcal{P}^n , est représenté par un vecteur de $(n + 1)$ éléments dont au moins un élément est non nul $P = [x_1, \dots, x_{n+1}]^t$. Les éléments de P sont appelés coordonnées homogènes du point. P est appelé vecteur de coordonnées homogènes. Deux vecteurs de dimension $(n + 1)$, $P = [x_1, \dots, x_{n+1}]^t$ et $P' = [x'_1, \dots, x'_{n+1}]^t$ représentent le même point si et seulement s'il existe un scalaire λ non nul tel que $P_i = \lambda P'_i$ pour $i = 1 \dots n + 1$. C'est ce qu'on appelle *égalité projective*. Ainsi la correspondance entre les points et les vecteurs de coordonnées n'est pas unique, mais définie à un facteur multiplicatif près.

Lorsque l'espace projectif est de dimension 1, 2 et $(n - 1)$, cet espace est appelé droite, plan et hyperplan. L'hyperplan d'équation $x_{n+1} = 0$ est appelé *hyperplan de l'infini*. Les points et les lignes se trouvant sur ce plan sont appelés *points à l'infini*.

et *lignes à l'infini* respectivement.

Dépendance linéaire

Les points P_1, \dots, P_m de l'espace projectif \mathcal{P}^n sont linéairement dépendants s'il existe un ensemble de scalaires $\lambda_1, \dots, \lambda_m$ non tous nuls tels que

$$\sum_{i=1}^m \lambda_i P_i = 0.$$

Base projective

Une base projective de \mathcal{P}^n est formée de $n+2$ points tel que tout sous-ensemble de $n+1$ points soit formé de points linéairement indépendants. La base projective canonique est formée par un ensemble de points $e_i = [0, \dots, 1, \dots, 0]^t$, $i = 1, \dots, n+1$, avec 1 pour le i^e élément. Tout point P de l'espace projectif peut être représenté comme une combinaison linéaire de n'importe quel sous-ensemble de $n+1$ points de la base. Par exemple le point P peut s'écrire

$$P = \sum_{i=1}^{n+1} \lambda_i e_i,$$

où les λ_i sont des scalaires.

2.1.2 Le plan projectif

Le plan projectif peut être considéré comme une généralisation du plan euclidien en introduisant deux nouvelles propriétés. Tout d'abord, la notion de distance est éliminée, afin d'introduire une nouvelle structure appelée plan affine qui est défini en fixant l'hyperplan de l'infini qui est formé par les points dont la dernière coordonnée est nulle (voir plus haut). La propriété la plus importante de ce plan est que le parallélisme est un invariant pour une transformation affine qui est représentée par une matrice de transformation homogène 3×3 caractérisée par une dernière colonne dont les éléments sont de la forme $[0, 0, 1]$. C'est une transformation qui tient compte de la composition d'une translation, une rotation, une projection parallèle et une mise à l'échelle. Une transformation affine applique une mise en échelle non-isotropique des coordonnées des points de l'espace projectif, ainsi par exemple un carré peut être transformé en un parallélogramme quelconque. En second lieu, ce modèle enlève le concept de lignes parallèles. Toute paire de lignes se croise en un point unique sous une projection perspective. Pour tenir compte de l'existence de lignes parallèles, c'est-à-dire se croisant à l'infini, la notion de *point idéal* ou *point à l'infini* est introduite (voir définition plus haut). Ainsi chaque ligne parallèle définit un *point idéal* différent. L'ensemble de ces points définit une ligne.

Le plan affine et la ligne formée par les points idéaux définissent le *plan projectif* (figure 2.1).

Ceci amène à énoncer deux axiomes qui sont à la base de la géométrie projective :

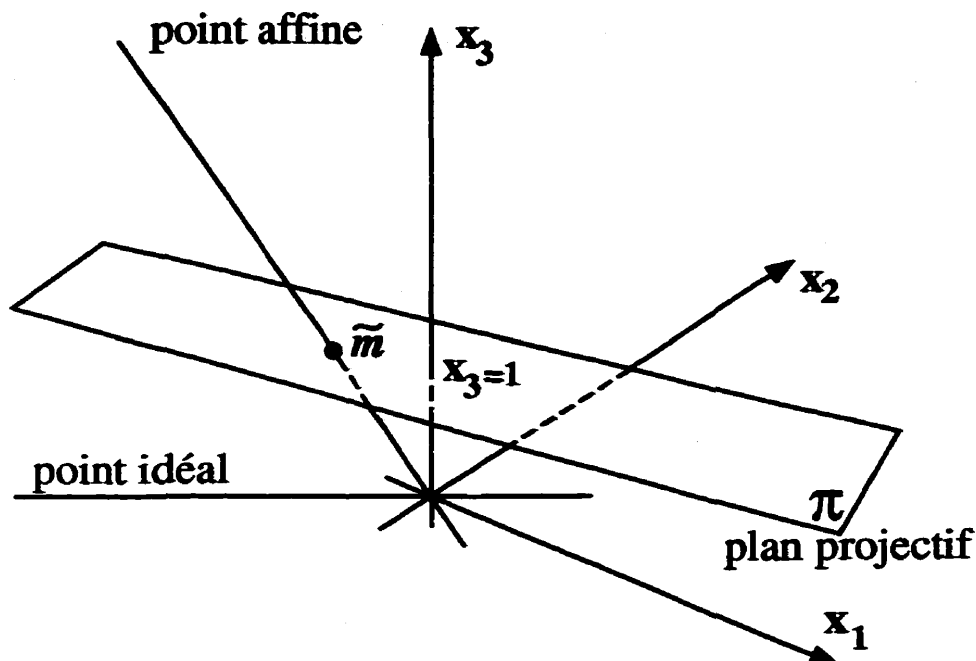


Figure 2.1 : Modèle de plan projectif

- A1 : Deux points quelconques déterminent une ligne unique.
- A2 : Deux lignes quelconques déterminent un point unique.

Ces deux axiomes introduisent une notion importante que nous allons utiliser par la suite, c'est la notion de : *dualité*. Ce concept est très important en géométrie projective, puisque tout résultat obtenu pour les points est valable pour les lignes et vice versa.

2.1.3 Coordonnées homogènes

En se basant sur le modèle projectif, un point m dans le plan projectif est représenté par trois coordonnées cartésiennes, $\overline{m} = [x_1, x_2, x_3]^t$. Tous les points sont définis à un

facteur multiplicatif près, ainsi $\bar{m} = [x_1, x_2, x_3]^t$ et $\bar{m}' = \lambda \bar{m} = [\lambda x_1, \lambda x_2, \lambda x_3]^t$ sont équivalents.

La relation avec les coordonnées cartésiennes conventionnelles dans le plan, (x, y) peut être établie en construisant un plan π_e , perpendiculaire à l'axe des x_3 et à une distance unité dans la direction x_3 . L'intersection du rayon passant par m et par le centre de projection donne un point, $\tilde{m} = [x, y, 1]^t$, appelé *représentation canonique* de m . La paire $[x, y]^t$ représente les coordonnées cartésiennes de m . Tout point $m = [x_1, x_2, x_3]^t$ tel que $x_3 = 0$ est appelé *point idéal* (voir figure 2.1) et l'ensemble de ces points définit une ligne appelée *ligne idéale*. Les coordonnées cartésiennes correspondantes au point projectif sont définies par

$$\tilde{m} = \left[\frac{x_1}{x_3}, \frac{x_2}{x_3}, 1 \right]^t = [x, y, 1]^t.$$

Il n'est pas nécessaire que le plan projectif soit perpendiculaire à x_3 , même si c'est la convention généralement utilisée.

2.1.4 La ligne projective

La représentation de la ligne dans le plan projectif est dérivée de la représentation analytique d'un plan passant par l'origine. L'équation de ce plan est donnée par

$$l_1 x_1 + l_2 x_2 + l_3 x_3 = 0. \quad (2.1)$$

Les coefficients $l = [l_1, l_2, l_3]^t$ correspondent aux coordonnées homogènes de la ligne projective. l et λl représentent la même ligne. Ici le cas $l_3 = 0$ correspond aux lignes passant par l'origine. La *ligne idéale* est donnée par $l = [0, 0, 1]^t$ qui a pour équation $x_3 = 0$. L'équation projective de la ligne peut s'écrire de plusieurs façons :

$$l \cdot m = l^t m = m^t l = 0.$$

La *dualité* entre les points et les lignes est indiquée par la forme symétrique de cette équation. Le rôle de l et de m peut être interchangé sans pour autant changer la forme de l'équation. La forme projective homogène d'une ligne peut être reliée à l'équation cartésienne standard de celle-ci. En coordonnées cartésiennes, l'équation d'une ligne s'écrit

$$n_x x + n_y y - d = 0,$$

où $n = (n_x, n_y)^t$ est la normale à la ligne et d la distance entre l'origine et la ligne en projection orthogonale à celle-ci. En comparant cette expression à l'équation homogène de la ligne (équation 2.1), on obtient la relation entre les paramètres cartésiens de la ligne et ses coefficients homogènes (figure 2.2). Les composantes normales sont données par

$$n_x = -d \frac{l_1}{l_3}, \quad n_y = -d \frac{l_2}{l_3}. \quad (2.2)$$

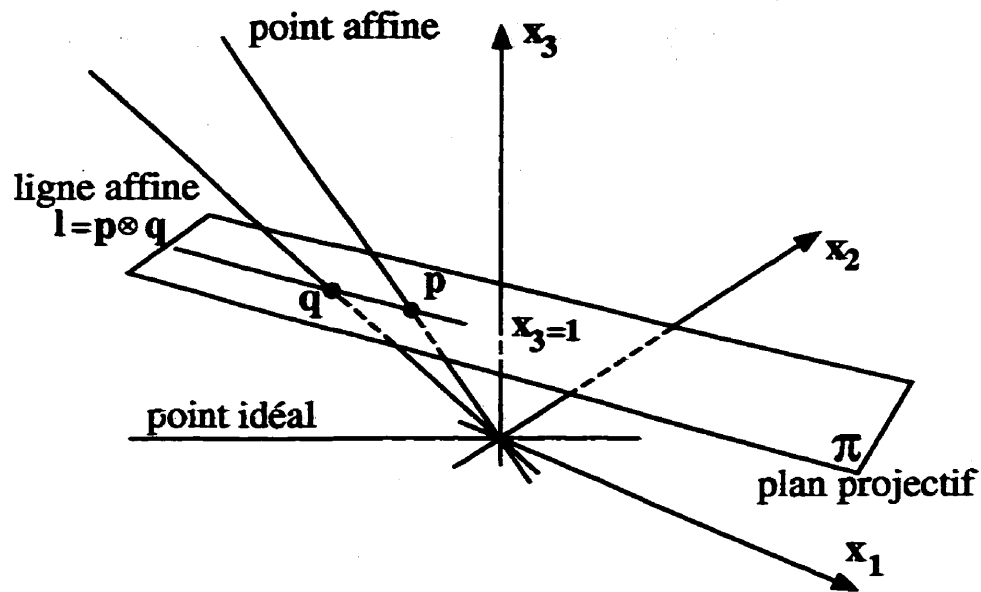


Figure 2.2 : Modèle d'obtention d'une ligne à partir de deux points dans le plan projectif

2.1.5 Transformations projectives

Une transformation projective ou colinéation est définie comme étant une transformation linéaire entre deux hyperplans projectifs. Une transformation projective de \mathcal{P}^n dans \mathcal{P}^m s'exprime sous la forme

$$\mathcal{P}^n \longrightarrow \mathcal{P}^m$$

$$P \longmapsto P' = T^t P.$$

Une transformation projective de \mathcal{P}^n dans lui-même est appelée *homographie*. L'ensemble des homographies forme un groupe appelé *groupe projectif*. L'homographie H entre deux plans projectifs est représentée par une matrice 3×3 avec 8 paramètres essentiels (figure 2.3). L'ensemble des transformations projectives est un sous-espace

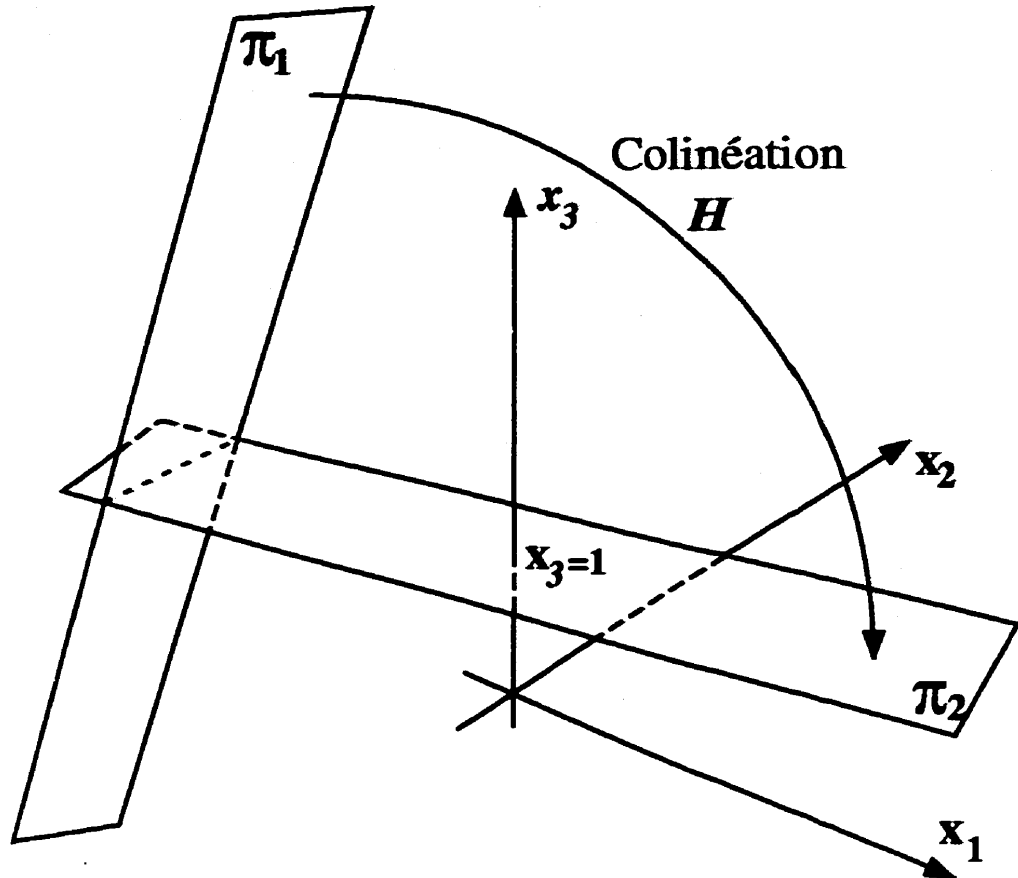


Figure 2.3 : Colinéation entre deux plans projectifs

de 8 dimensions de l'espace de 9 dimensions formé par les éléments de la matrice H .

La transformation d'un plan projectif π_1 vers un autre plan projectif π_2 , est représentée par

$$\begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad (2.3)$$

où

$$m' = Hm.$$

La représentation de la transformation en coordonnées cartésiennes fait apparaître la nature non linéaire de la transformation projective dans un espace euclidien ou affine :

$$x' = \frac{x'_1}{x'_3} = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}}, \quad (2.4)$$

$$y' = \frac{x'_2}{x'_3} = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}}. \quad (2.5)$$

2.1.6 Transformation projective de lignes

La dualité entre les points et les lignes dans un plan projectif implique que la transformation projective de lignes est linéaire aussi. Soit l'équation d'un point incident à une ligne,

$$l_1x_1 + l_2x_2 + l_3x_3 = 0,$$

$$l^t m = 0.$$

Etant donnée la transformation $m' = Hm$ pour un point, la transformation inverse est donnée par $m = H^{-1}m'$. En substituant cette transformation inverse dans l'équation de la ligne on obtient

$$l^t H^{-1}m' = 0.$$

Les transformations projectives préservent la colinéarité des points. Donc l'équation de la ligne transformée est donnée par $l'^t m'$. Les coordonnées de la ligne l' sont

données par

$$l' = [H^{-1}]^t l.$$

Si on pose $H_t = H^t$, on obtient alors

$$l' = H_t^{-1} l. \quad (2.6)$$

Ainsi, les lignes dans le plan projectif se transforment linéairement, comme les points, mais avec une matrice de transformation qui est la transposée de l'inverse de la matrice qui définit la transformation des points.

2.1.7 Calcul de l'homographie

La matrice de transformation projective H , nécessite 8 paramètres indépendants pour définir une application unique. Comme chaque paire de points appariés donne deux équations en coordonnées cartésiennes, il est donc nécessaire de trouver au moins 4 correspondances entre deux plans transformés projectivement pour définir la matrice de transformation de façon unique.

Solution linéaire non homogène

Le système étant résolu à un facteur multiplicatif près, on peut donc choisir $h_{33} = 1$.

Ainsi, pour quatre paires de points correspondants on a

$$[x'_i, y'_i, 1]^t = \lambda H [x_i, y_i, 1]^t.$$

Ici λ représente le facteur d'échelle, puisque H est calculé à un facteur multiplicatif près. Le système d'équations linéaires résultant assure l'unicité de la solution à partir de quatre paires de points correspondants, à condition qu'aucun des points ne soit colinéaire à deux autres. Si on a plus de quatre points correspondants on peut utiliser une procédure de minimisation pour résoudre le système surdéterminé.

La méthode décrite nous donne une solution linéaire non homogène. Son principal désavantage est qu'on obtient une mauvaise estimation des éléments de H si l'élément qu'on a fixé à 1 est nul en réalité.

Solution homogène

Pour éviter le problème rencontré avec la solution non homogène, on utilise une approche basée sur la décomposition en valeurs singulières (SVD) (Golub et Van Loan, 1989). C'est la meilleure approche du point de vue algorithmique. Plus loin dans ce chapitre (voir section 2.4.1.4), nous verrons comment nous pourrions résoudre un système linéaire en utilisant cette décomposition.

En posant les éléments de la matrice H sous la forme d'un vecteur

$h = [h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33}]^t$ et en utilisant les relations 2.4 et 2.5, l'équation homogène 2.3 pour n points devient

$$Ah = 0,$$

où A est la matrice $2n \times 9$ suivante :

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1x'_1 & -y_1x'_1 & -x'_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1y'_1 & -y_1y'_1 & -y'_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2x'_2 & -y_2x'_2 & -x'_2 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2y'_2 & -y_2y'_2 & -y'_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & 1 & 0 & 0 & 0 & -x_nx'_n & -y_nx'_n & -x'_n \\ 0 & 0 & 0 & x_n & y_n & 1 & -x_ny'_n & -y_ny'_n & -y'_n \end{bmatrix}.$$

Ici on utilise une normalisation différente de celle mentionnée précédemment. On fixe la norme du vecteur h égale à 1. La solution qui minimise la norme de Fobenius¹ $\| Ah \|$ avec la contrainte $\| h \| = 1$, est donnée par le vecteur propre correspondant à la plus petite valeur propre de $A^t A$. Ce vecteur propre est obtenu directement à partir de la décomposition en valeurs singulières de A (voir section 2.4.1.4).

¹Ici $\| \cdot \|$ représente la norme de Fobenius. Pour un vecteur donné $a = (a_1, \dots, a_n)$, on a : $\| a \| = \sqrt{a \cdot a} = \sqrt{a_1^2 + \dots + a_n^2}$.

Solution géométrique non linéaire

Dans ce cas la matrice H est obtenue en minimisant la somme des distances euclidiennes, dénotées par d , entre les points mesurés et les points reprojetés par H . Soit m_e un point dans le premier plan image et m'_e son correspondant dans le deuxième plan image, la transformation H qui projète m_e dans le deuxième plan image doit minimiser la distance entre le point projeté Hm_e et le point correspondant m'_e . De même elle doit minimiser la distance de la projection inverse puisque H est inversible. Cette minimisation s'écrit

$$\min_h \sum_i (d^2(m'_e, Hm_e) + d^2(m_e, H^{-1}m'_e)).$$

Homographie à partir des lignes

Comme la ligne est la duale du point, on peut utiliser les techniques décrites plus haut pour résoudre l'homographie à partir d'une correspondance de lignes. Il suffit de remplacer les points par les lignes dans les équations précédentes et d'utiliser la relation d'homographie liant deux lignes appariées (équation 2.6).

2.2 Modèle de la caméra

2.2.1 Modèle projectif

Comme la transformation projective contient les différents modèles de transformations utilisés en vision, elle représente donc le modèle le plus général à partir duquel on peut retrouver les autres modèles.

Les propriétés géométriques essentielles d'une projection pour former une image peuvent être modélisées par une application d'un espace tridimensionnel projectif vers un plan projectif, qu'on peut représenter par une simple transformation linéaire homogène (Faugeras, 1993).

Un point de l'espace projectif 3D est représenté par un vecteur de coordonnées homogènes de dimension 4. Une transformation projective générale est alors définie par une matrice 4×4 :

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \\ t_{41} & t_{42} & t_{43} & t_{44} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}.$$

Une projection vers un espace de dimension inférieure est obtenue en éliminant un des coordonnées de l'espace projectif transformé. Nous choisissons, par exemple, le plan défini par $x_4 = 0$. Ainsi tous les points sur le plan peuvent être représentés

par un vecteur de coordonnées homogènes, $[x_1, x_2, x_3]^t$. Le choix de $x_4 = 0$ est général, puisqu'on peut toujours transformer tout plan en celui-ci par une simple transformation projective de l'espace 3D. La projection de l'image est donnée par

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \quad (2.7)$$

ou bien

$$m = T M.$$

La matrice de transformation homogène T est estimée à un facteur multiplicatif près ; elle a 11 paramètres indépendants. Six points au moins dans l'espace et leurs points images correspondants permettent le calcul de T . Les techniques utilisées sont les mêmes que pour l'estimation de l'homographie (voir section 2.1.7).

2.2.2 Modèle de projection perspective

L'application définie par l'équation 2.7 tient compte de plusieurs aspects dans la formation d'images, par exemple les effets des distorsions introduites par les différentes sources de bruit (distorsions verticales et horizontales causées par la lentille de la caméra, erreur sur la position du centre optique, etc.). La matrice T peut être mise sous une forme qui ne prend en compte que la projection d'un espace 3D dans un

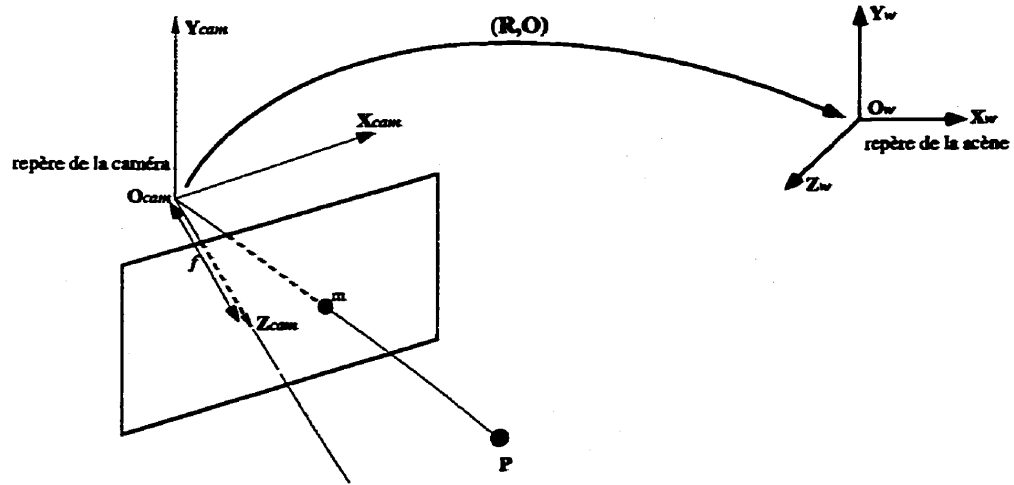


Figure 2.4 : Modèle de projection perspective

plan image à partir d'un point central (projection centrale). Ce modèle est le modèle de projection perspective de la caméra (Faugeras, 1993). La géométrie de la caméra perspective est définie à la figure 2.4.

Ainsi la transformation euclidienne d'un point $P = [X, Y, Z]$ du repère de la scène au repère de la caméra est donnée par

$$P_{cam} = R(P_w - O),$$

où la matrice

$$R = \begin{bmatrix} R_1 \\ R_2 \\ R_3 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

définit la rotation entre le repère de la scène et le repère de la caméra, le vecteur $R_i = [r_{i1}, r_{i2}, r_{i3}]$, $i = 1, 2, 3$ représente la ligne i de la matrice R , $O = [o_1, o_2, o_3]^t$ est

le vecteur de translation entre l'origine du repère de la scène et l'origine du repère de la caméra, $P_{cam} = [X_{cam}, Y_{cam}, Z_{cam}]^t$ est le point P représenté par rapport au repère de la caméra et $P_w = [X_w, Y_w, Z_w]^t$ représente le point P dans le repère de la scène. L'origine de la caméra est confondue avec le centre de projection. La transformation appliquée est donc la composition d'une translation O suivie d'une rotation R .

Ces deux transformations sont représentées, en notation matricielle, par une matrice de transformation homogène :

$$T_{ext} = \begin{bmatrix} R_1 & -(R_1 \cdot O) \\ R_2 & -(R_2 \cdot O) \\ R_3 & -(R_3 \cdot O) \\ 0 & 1 \end{bmatrix}.$$

Le point est projeté sur le plan image par la matrice

$$T_{proj} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 0 \end{bmatrix},$$

où f est la distance focale de la caméra. La matrice de transformation composée,

$T = T_{proj} T_{ext}$ est donnée par

$$T = \begin{bmatrix} R_1 & -(R_1 \cdot O) \\ R_2 & -(R_2 \cdot O) \\ R_3/f & -(R_3 \cdot O)/f \end{bmatrix}. \quad (2.8)$$

Le cas le plus fréquent en vision est celui où le centre de projection est à l'origine du repère de la scène et les axes de la caméra sont alignés avec les axes du même repère.

Dans ce cas $T = T_{proj}$. Ce qui nous amène aux équations de projection perspective utilisées en vision. Pour un point de l'espace 3D, noté $[X, Y, Z, 1]^t$, on a

$$x = \frac{x_1}{x_3} = f \frac{X}{Z},$$

$$y = \frac{x_2}{x_3} = f \frac{Y}{Z}.$$

Paramètres intrinsèques et paramètres extrinsèques de la caméra

L'équation 2.8 peut être réécrite pour tenir compte des paramètres qui définissent la géométrie interne de la caméra. Le modèle analytique de la caméra devient la composition d'une matrice de paramètres intrinsèques T_{int} et d'une matrice qui définit l'orientation et la position externe du repère de la caméra par rapport au repère de

la scène, T_{ext} :

$$T = T_{int} T_{ext} = \begin{bmatrix} s_x & 0 & u_o \\ 0 & s_y & v_o \\ 0 & 0 & 1/f \end{bmatrix} \begin{bmatrix} R_1 & -(R_1 \cdot O) \\ R_2 & -(R_2 \cdot O) \\ R_3 & -(R_3 \cdot O) \end{bmatrix}, \quad (2.9)$$

où s_x , s_y représentent les tailles des pixels suivant x et y respectivement, u_o et v_o désignent la position du point principal, c'est-à-dire l'intersection du plan image et de l'axe des z . La matrice T définit la projection perspective. C'est une matrice 3×4 , appelée *matrice de projection perspective* ; elle projette un point 3D de la scène, $[X, Y, Z, 1]^t$, dans le plan image :

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = T_{int} T_{ext} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = T \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (2.10)$$

En général on choisit le centre de projection à l'origine du repère de la scène et les axes de la caméra alignés avec les axes de celui-ci. Ainsi dans un système de

coordonnées normalisé où $f = 1$, l'équation 2.10 s'écrit

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & u_o \\ 0 & s_y & v_o \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}.$$

2.3 Caméras stéréoscopiques

Le modèle stéréoscopique est représenté par le déplacement entre deux repères reliés à chacune des deux caméras de la paire stéréoscopique, caméra gauche et caméra droite. En général on choisit le repère d'une des caméras, par exemple la caméra gauche, comme étant le repère de la scène. Ainsi le modèle stéréoscopique est défini par une transformation homogène entre le repère de la caméra gauche et le repère de la caméra droite (Faugeras, 1993; Horn, 1986). La position et l'orientation de cette dernière sont représentées par une matrice homogène 4×4 ,

$$\overline{P} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix},$$

où R est la matrice de rotation et t est le vecteur translation. Ce sont les paramètres extrinsèques du modèle stéréoscopique.

La figure 2.5 décrit la géométrie entre deux vues perspectives représentant une paire stéréoscopique. La ligne qui relie les centres de projection des deux caméras,

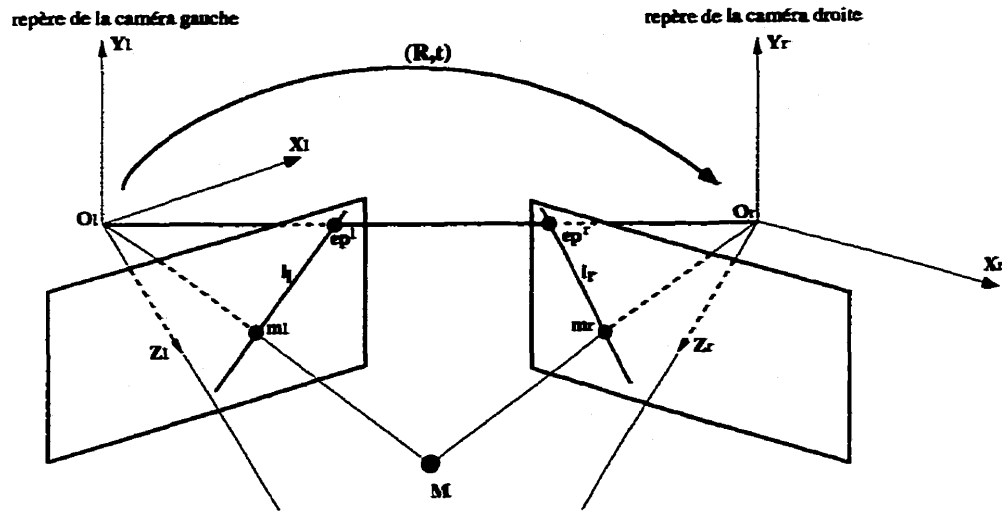


Figure 2.5 : Modèle de caméras stéréoscopiques

$t = \overline{O_l O_r}$, intersecte chaque plan image en un point appelé *point épipolaire* ou *épipole*.

Pour un point de la scène, M , le plan $(O_l O_r M)$, coupe chaque plan image en une ligne appelée *ligne épipolaire*. Cette ligne représente la projection d'une ligne 3D sur le plan image. On peut observer sur la figure 2.5 que la projection m_l d'un point M de l'espace sur l'image gauche et le centre de projection de la caméra gauche O_l définissent un rayon $O_l m_l$. Le point M est situé sur ce rayon et l'image de ce rayon sur le plan image droit constitue la ligne épipolaire l_r (voir figure 2.5) .

Soit M , un point visible de la scène. Ses coordonnées par rapport aux deux vues sont respectivement $M_l = [X_l, Y_l, Z_l]^t$ pour le repère de la caméra gauche et $M_r = [X_r, Y_r, Z_r]$ pour le repère de la caméra droite. Ces deux points sont reliés par la transformation suivante :

$$M_r = R M_l + t. \quad (2.11)$$

On peut constater sur la figure 2.5 que deux rayons correspondants, passant par

deux centres de projections, ont pour intersection un même point de l'espace 3D. En général on peut écrire cette relation sous la forme,

$$M_l R(t \otimes M_r) = 0 \quad (2.12)$$

où le produit vectoriel \otimes par t correspond à une multiplication par une matrice antisymétrique du vecteur translation de la forme

$$\Omega(t) = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix}.$$

L'équation 2.12 devient alors

$$\begin{aligned} M_l^t \Omega(t) R M_r &= 0, \\ M_l^t E M_r &= 0, \end{aligned} \quad (2.13)$$

où $E = \Omega(t)R$ est appelée *matrice essentielle* (Longuet-Higgins, 1981; Hartley, 1992b; Hartley, 1992a). Une des propriétés les plus importantes de cette matrice est qu'elle est de rang égal à 2 (Faugeras, 1993).

Les coordonnées images sont proportionnelles aux coordonnées cartésiennes de

l'espace 3D :

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \frac{1}{Z_i} \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} \quad i = l, r.$$

On a la contrainte suivante en coordonnées homogènes entre les points appariés des deux plans images :

$$m_l^t E m_r = 0. \quad (2.14)$$

La même équation peut être étendue au cas d'une transformation projective quelconque. Une caméra projective peut être formée en appliquant une transformation projective planaire à une image perspective. Appliquons à chaque image une transformation planaire projective, $m_l = T_l m'_l$, $m_r = T_r m'_r$. L'équation 2.14 devient

$$m_l^{t'} T_l^t E T_r m'_r = m_l^{t'} Q m'_r = 0, \quad (2.15)$$

où $Q = T_l^t E T_r$ représente la matrice essentielle pour une transformation projective quelconque. Cette généralisation présente un intérêt particulier lorsqu'on introduit les paramètres intrinsèques des caméras dans le cas où elles sont non calibrées.

Géométrie épipolaire et paramètres intrinsèques

Pour tenir compte des paramètres qui définissent la géométrie interne de la caméra, on introduit la matrice de paramètres intrinsèques dans l'équation 2.14. On obtient

alors l'équation² de transformation épipolaire de la même forme que l'équation 2.15 :

$$m_i^t (A_i^t)^{-1} \Omega(t) R A_r^{-1} m_r = 0. \quad (2.16)$$

En posant

$$F = (A_i^t)^{-1} \Omega(t) R A_r^{-1},$$

l'équation 2.16 devient,

$$m_i^t F m_r = 0. \quad (2.17)$$

F est appelée *matrice fondamentale* (Faugeras, 1993; Hartley, 1994a). Comme la matrice essentielle E (voir section 2.3), F est de rang 2.

Formulation linéaire de la contrainte épipolaire

Ainsi, pour une paire de points images appariés, $m_l = [x_l, y_l, 1]$ et $m_r = [x_r, y_r, 1]^t$, l'équation épipolaire 2.17 s'écrit,

$$x_r x_l f_{11} + x_r y_l f_{21} + x_r f_{31} + y_r x_l f_{12} + y_r y_l f_{22} + y_r f_{32} + x_l f_{13} + y_l f_{23} + f_{33} = 0. \quad (2.18)$$

Cette équation peut être écrite sous une forme linéaire :

$$u_i^t f = 0, \quad (2.19)$$

²Ici on a gardé la même notation que l'équation 2.14 pour représenter les points, m_i .

avec

$$u_i = [x_l x_r, x_l y_r, x_l, y_l x_r, y_l y_r, y_l, x_r, y_r, 1]^t,$$

$$f = [f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33}]^t,$$

où les f_{ij} représentent les éléments de la matrice fondamentale F .

Pour un ensemble de n points appariés, les n équations linéaires de la forme de l'équation 2.19 donnent un système d'équations linéaires :

$$U_n f = 0,$$

où $U_n = [u_1, \dots, u_n]^t$.

Ce système d'équations épipolaires permet l'estimation de la géométrie épipolaire de la paire de caméras stéréoscopiques par le calcul de la matrice fondamentale F . La matrice F permet, à partir d'un point sur la vue de gauche m_l , de calculer sa ligne épipolaire sur la vue de droite l_r . Cette ligne épipolaire passe par le point image m_r qui est le correspondant du point image m_l sur l'image de droite. Toutes les lignes épipolaires se joignent en un point ep appelé point épipolaire. Le plan Π , qui passe par les deux points appariés m_l et m_r , leurs lignes épipolaires respectives l_r et l_l et le point M de l'espace 3D qui leur correspond est appelé le *plan épipolaire* (figure 2.6).

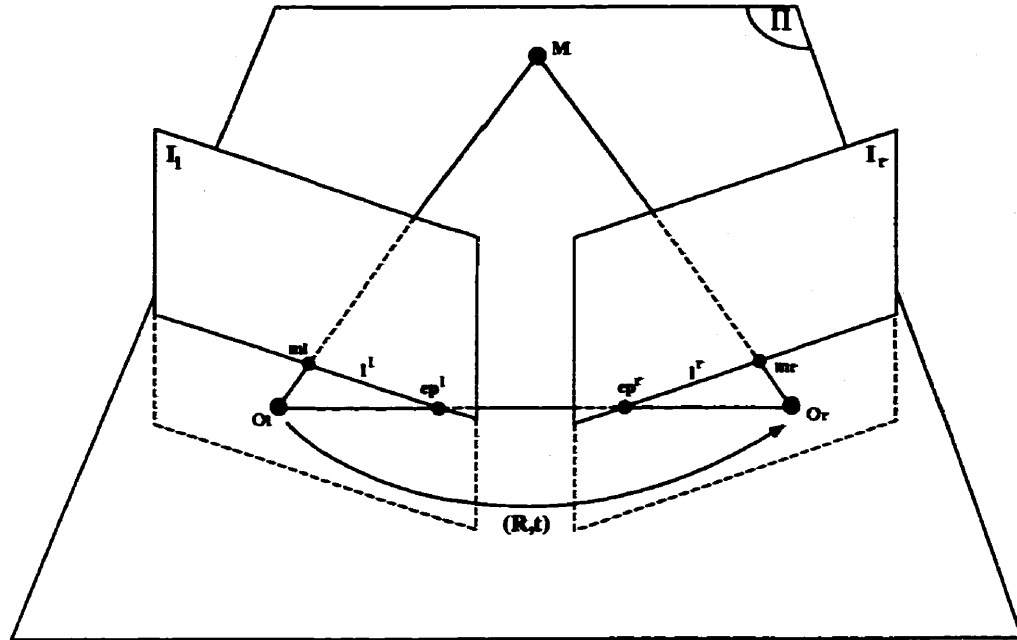


Figure 2.6 : Géométrie épipolaire

2.4 Calcul de la matrice fondamentale

Depuis que Longuet-Higgins (Longuet-Higgins, 1981) a introduit une méthode linéaire pour le calcul de la matrice essentielle E , appelée *l'algorithme des huit points*, afin d'estimer la structure de la scène, de nombreux chercheurs se sont penchés sur le problème du calcul de la matrice essentielle dans le cas de caméras calibrées, c'est-à-dire dont on connaît les paramètres intrinsèques a priori. Par la suite, le problème d'estimation de la matrice fondamentale F , dans le cas de caméras non calibrées, a attiré l'attention de plusieurs chercheurs (Faugeras et al., 1992; Hartley, 1993; Deriche et al., 1994; Csurka et al., 1997; Torr, Zisserman et Maybank, 1998; Zhang, 1997; Zhang, 1998c; Vieville, Faugeras et Luong, 1996). Vu l'importance de F dans la définition de la géométrie épipolaire d'une paire de caméras stéréoscopiques non

calibrées, beaucoup d'efforts ont été déployés pour trouver de nouvelles techniques pour la calculer. Les mêmes techniques précédemment développées pour le calcul de la matrice essentielle E ont été par la suite utilisées pour le calcul de la matrice fondamentale F . Toutefois, vu la sensibilité du calcul de cette dernière aux erreurs d'appariements, de plus en plus de méthodes dites robustes ont été développées pour palier cet inconvénient. Comme la matrice essentielle E , la matrice fondamentale F peut être utilisée pour : la reconstruction de la scène (Hartley, 1994c; Zhang, Luong et Faugeras, 1996; Laveau et Faugeras, 1996; Hartley, 1995; Horaud et Csurka, 1998; Boufama et Mohr, 1995), à une transformation projective près, à partir de deux vues non calibrées, la rectification de l'image (Faugeras, 1993; Seitz et Dyer, 1996b), le calcul d'invariants projectifs (Shashua, 1994a; Mundy, Zisserman et Forsyth, 1993; Hartley, 1994b), la détection de fausses mesures et la correspondance stéréo (Zhang, Deriche, Faugeras et Luong, 1995). Dans la gamme des techniques pour calculer F on distingue principalement deux classes, les méthodes linéaires et les méthodes non linéaires (Zhang, 1996; Torr et Murray, 1997).

2.4.1 Méthodes Linéaires

2.4.1.1 Algorithme des huit points

L'équation de base définissant la matrice fondamentale, équation 2.17, donne lieu à une équation linéaire et homogène en termes des neuf coefficients de la matrice fondamentale F , équation 2.19, comme nous avons vu plus haut. Comme la matrice

fondamentale est déterminée à un facteur d'échelle près, elle a donc huit paramètres indépendants. Ainsi, si nous disposons de huit appariements, nous pouvons déterminer une solution unique pour F , définie à un facteur d'échelle près, en résolvant un système linéaire composé de huit équations 2.19. Cette approche fut introduite par Longuet-Higgins (Longuet-Higgins, 1981) puis par Tsai et Huang (Tsai, 1987), et constitue ce qu'on a appelé l'algorithme des huit points. Elle a été largement étudiée dans la littérature pour la détermination de la matrice essentielle et du mouvement relatif entre scène et caméras (Mitiche, 1994). Un grand nombre d'algorithmes développés par la suite en sont dérivés. Plusieurs études sur la sensibilité au bruit de cet algorithme ont été menées et ont démontrées qu'il était en effet très sensible. Cet algorithme a été généralisé pour calculer la géométrie épipolaire à partir de caméras non calibrées.

2.4.1.2 Définition du critère linéaire

Pour résoudre le système d'équations épipolaires, nous avons besoin de huit points appariés. En pratique nous disposons d'un nombre n plus grand d'appariements. Nous utilisons une méthode des moindres carrés pour résoudre par rapport à F :

$$\hat{f} = \min_F \sum_i (m_i^t F m_r)^2, \quad (2.20)$$

qui s'écrit aussi sous la forme

$$\hat{f} = \min_f \| U_n f \|^2. \quad (2.21)$$

Le vecteur f est défini à un facteur d'échelle près. Une solution triviale existe pour le problème de minimisation 2.21 : c'est $f = 0$, qui n'est pas la solution que nous recherchons. Pour résoudre ce problème, il faut imposer certaines contraintes sur les coefficients de la matrice fondamentale.

2.4.1.3 Technique des moindres carrés linéaire

Cette méthode consiste à fixer un des coefficients de la matrice fondamentale F égale à 1, puis à résoudre, par rapport aux huit autres coefficients, l'équation 2.21 en utilisant une technique de moindres carrés linéaire. L'inconvénient de cette approche est que nous ne disposons pas d'information a priori sur la matrice fondamentale. Donc, si on choisit un des coefficients à fixer à 1 qui en réalité est égal à *zéro*, la matrice obtenue sera très différente de la vraie matrice fondamentale. En général, le coefficient qui est choisi égal à 1 est f_{33} . Toutefois, par exemple dans le cas d'une translation pure, la matrice fondamentale est antisymétrique et en conséquence le dernier coefficient de la matrice est égal à 0. Il faut donc éviter alors de le normaliser à 1. Pour remédier à cet inconvénient, nous pourrions essayer toutes les neuf possibilités, en normalisant à chaque fois un des coefficients à 1, puis de résoudre l'équation 2.21 et de choisir la meilleure estimation obtenue, c'est-à-dire celle pour laquelle l'équation 2.21 est

minimisée.

2.4.1.4 Minimisation sous contrainte

La seconde méthode consiste à imposer une contrainte sur la norme de f . Nous pouvons fixer cette norme à 1, $\|f\| = 1$. Ainsi, tous les coefficients de f contribuent d'une manière équivalente au processus de minimisation. La relation 2.21 devient alors :

$$\hat{f} = \min_f \|U_n f\|^2, \quad \text{avec } \|f\| = 1. \quad (2.22)$$

C'est un problème de minimisation sous contraintes. Il peut être transformé en un problème de minimisation sans contraintes en introduisant les multiplicateurs de Lagrange (Golub et Van Loan, 1989). L'équation 2.22 s'écrit alors :

$$\hat{f} = \min_f \mathcal{F}(f, \lambda), \quad (2.23)$$

avec

$$\mathcal{F}(f, \lambda) = \|U_n f\|^2 + \lambda(1 - \|f\|^2), \quad (2.24)$$

où λ est le multiplicateur de Lagrange. En posant la dérivée première de $\mathcal{F}(f, \lambda)$ par rapport à f égale à zéro, nous obtenons la relation suivante :

$$U_n^t U_n f = \lambda f. \quad (2.25)$$

La solution f est donc le vecteur propre de la matrice 9×9 , $U_n^t U_n$, et λ est la valeur propre correspondant à ce vecteur. Comme $U_n^t U_n$ est symétrique et semi-définie positive, toutes ses valeurs propres sont réelles et positives ou bien nulles. Soient les valeurs propres de $U_n^t U_n$ prises dans un ordre décroissant:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_9 \geq 0.$$

Il y a donc 9 solutions possibles : $\lambda = \lambda_i$, $i = 1, \dots, 9$. En substituant la solution dans 2.24 nous obtenons

$$\mathcal{F}(f, \lambda_i) = \lambda_i.$$

Comme nous cherchons à minimiser $\mathcal{F}(f, \lambda)$, la solution à l'équation 2.22 est donc le vecteur propre de la matrice $U_n^t U_n$ correspondant à la plus petite valeur propre soit λ_9 .

2.4.1.5 Contrainte de rang 2

La formulation linéaire du problème présente l'avantage de permettre une solution analytique simple. Toutefois, elle reste très sensible au bruit. Une des raisons est que la contrainte de rang 2 (voir section 2.3) n'est pas satisfaite. En effet, il est bien connu que la matrice fondamentale est une matrice 3×3 , mais que son rang est égal à 2 ($\det F = 0$). Une des techniques utilisées consiste à imposer la contrainte de rang 2 *a posteriori* (Hartley, 1992b). La matrice fondamentale F , calculée par l'une

des méthodes mentionnées plus haut, est remplacée par la matrice \hat{F} qui minimise la norme de Fobenius $\|F - \hat{F}\|$ sous la contrainte $\det F = 0$,

$$\hat{f} = \min_F \|F - \hat{F}\|, \quad \text{avec } (\det F = 0). \quad (2.26)$$

Pour résoudre 2.26, posons

$$F = USV^t,$$

ce qui représente la décomposition en valeurs singulières de la matrice F , où $S = \text{diag}(\sigma_1, \sigma_2, \sigma_3)$ est la matrice diagonale telle que $\sigma_1 \geq \sigma_2 \geq \sigma_3$ (avec σ_i la i^e valeur singulière). U et V sont deux matrices orthogonales. Il existe donc une matrice \hat{F} telle que

$$\hat{F} = U\hat{S}V^t$$

qui minimise la norme de Fobenius de $F - \hat{F}$, avec $\hat{S} = \text{diag}(\sigma_1, \sigma_2, 0)$.

2.4.1.6 Minimisation d'un critère géométrique

La forme de la relation 2.21 utilise un critère de minimisation qui n'exploite pas l'information contenue dans la géométrie épipolaire de l'image. Un critère valide serait celui qui essaierait de minimiser une mesure ayant une *signification physique*. La mesure qui est proposée dans la littérature (Luong et Faugeras, 1992; Luong et Vieville, 1994; Zhang, 1998b) est la distance entre les points appariés et les lignes

épipolaires qui leur correspondent,

$$\min_F \sum_i d^2(m_i, Fm_r), \quad (2.27)$$

où m_r est un point du plan image gauche, et $l_i = Fm_r$ est la ligne épipolaire qui lui correspond sur le plan image droit. Idéalement la ligne épipolaire de gauche, $l_i = Fm_r = [l_1, l_2, l_3]^t$, passe par un point m_i se trouvant sur l'image gauche qui est le correspondant du point m_r . En pratique cette ligne épipolaire se trouve à une distance d de ce point :

$$d(m_i, l_i) = \frac{m_i^t l_i}{\sqrt{l_1^2 + l_2^2}} = \frac{1}{c} m_i^t F m_r, \quad (2.28)$$

avec $c = \sqrt{l_1^2 + l_2^2}$.

La relation 2.27 peut être minimisée en F , en utilisant une procédure itérative de moindres carrés pondérés.

Relation avec le critère linéaire

En utilisant la distance entre les points et les lignes épipolaires 2.28, le critère linéaire, équation 2.20, devient

$$\min_F \sum_{i=1}^n c_i^2 d^2(m_i, l_i).$$

Le critère linéaire tend à minimiser non seulement une mesure physique $d(m_i, l_i)$, mais aussi c_i , qui n'est pas une mesure physique par rapport à la géométrie épipolaire

entre les deux vues. Ainsi, le critère linéaire introduit un biais dans le calcul de la matrice fondamentale et tend à placer l'épipole au centre de l'image dans certains cas (les lignes épipolaires vont se croiser en un point qui se trouve proche du centre de l'image tel que observé par Luong (Luong et Faugeras, 1994b; Luong et Faugeras, 1994a)).

2.4.1.7 Normalisation des données

L'analyse, du point de vue numérique, de la grande instabilité de la méthode linéaire lorsque les coordonnées, en pixels, sont directement utilisées dans la procédure de minimisation a amené la proposition de méthodes permettant le traitement des données à l'entrée avant de procéder à la minimisation du critère linéaire pour le calcul de la matrice fondamentale. Si nous prenons la matrice de données utilisée en 2.25, $U_n^t U_n$, sa décomposition en valeurs singulières s'écrit

$$U_n^t U_n = U D U^t,$$

où U est une matrice orthogonale, et $D = \text{diag}(\lambda_1, \dots, \lambda_9)$ avec les λ_i en ordre décroissant. $\kappa = \lambda_1/\lambda_9$ représente la *condition* de la matrice $U_n^t U_n$. La condition représente un très bon indice d'analyse de stabilité des problèmes linéaires. Si κ est grand, un peu de bruit en entrée peut entraîner une grande variation des résultats en sortie. Dans notre cas la matrice $U_n^t U_n$ est mal conditionnée, sa condition, κ , est très grande ($\sim 10^4$). Ceci est dû à la non-homogénéité des coordonnées images utilisées.

Plusieurs méthodes ont été développées pour résoudre des systèmes mal conditionnés, la plupart utilisent une normalisation de la matrice de données qui tend à réduire la condition de cette dernière. Une technique de normalisation des coordonnées images fut proposée par Hartley (Hartley, 1997a) pour bien conditionner la matrice $U_n^t U_n$.

Supposons qu'on remplace les coordonnées des points m_{li} , de l'image gauche par $\widehat{m}_{li} = T m_{li}$, et les coordonnées des points m_{ri} , de l'image droite par $\widehat{m}_{ri} = T' m_{ri}$, avec T et T' deux matrices 3×3 . En remplaçant dans $m_{li}^t F m_{ri} = 0$, nous obtenons :

$$\widehat{m}_{li}^t T^{-t} F T'^{-1} \widehat{m}_{ri} = 0.$$

Cette relation implique que $\widehat{F} = T^{-t} F T'^{-1}$ représente la matrice fondamentale correspondant aux appariements $\widehat{m}_{li} \leftrightarrow \widehat{m}_{ri}$. La nouvelle méthode d'estimation de la matrice fondamentale est résumée ci-dessous :

1. Transformer les coordonnées image par T et T' , tels que $\widehat{m}_{li} = T m_{li}$ et $\widehat{m}_{ri} = T' m_{ri}$.
2. Trouver la matrice fondamentale \widehat{F} qui correspond à $\widehat{m}_{li} \leftrightarrow \widehat{m}_{ri}$ par l'une des méthodes décrites plus haut.
3. Retrouver la matrice fondamentale originale par $F = T^t \widehat{F} T'$.

Il faut donc choisir deux transformations T et T' qui permettent une normalisation efficace des données dans le sens de la réduction de la condition de la matrice $U_n^t U_n$.

L'analyse montre qu'une normalisation des données telle que leur moyenne soit de l'ordre de l'unité peut améliorer significativement les résultats de la méthode linéaire. Deux techniques peuvent être utilisées pour faire cette normalisation. La première consiste à centrer l'origine de l'image, généralement prise en haut à gauche, puis normaliser les coordonnées, de manière à avoir des points à une distance moyenne égale à $\sqrt{2}$ par rapport à l'origine. Ces transformations sont menées indépendamment sur les deux images, cette technique est appelée *normalisation isotropique*. La deuxième technique procède de manière équivalente, d'abord en amenant le centroïde des points à l'origine, puis en effectuant une normalisation sur l'ensemble des points pour que les deux moments soient égaux à 1. Ceci se fait d'une manière simple en une seule opération, en utilisant la décomposition de Cholesky. Les points m_i sont utilisés pour former une matrice $\sum_i m_i m_i^t$. En la décomposant par la transformation de Cholesky, nous obtenons,

$$\sum_i m_i m_i^t = N K K^t,$$

où K est une matrice triangulaire supérieure. La transformation choisie est : $\widehat{m}_i = K^{-1} m_i$. Les points \widehat{m}_i sont tels que

$$\sum_i \widehat{m}_i \widehat{m}_i^t = \sum_i K^{-1} m_i m_i^t K^t = N I,$$

où I est la matrice identité. Les points \widehat{m}_i ont donc leur centroïde à l'origine et les deux moments principaux sont égaux à 1. Comme K est une matrice triangulaire

supérieure, elle représente une transformation affine, c'est pourquoi on appelle cette technique, *normalisation affine* ou *normalisation non-isotropique*.

2.4.2 Méthodes non linéaires

Nous avons vu précédemment qu'un nouveau critère a été introduit dans la littérature (Luong et Faugeras, 1994b). Il est basé sur la minimisation de la distance entre les points et les lignes épipolaires correspondantes (équation 2.27). L'extension de ce critère de minimisation aux deux images formant la paire stéréoscopique, nous donne la relation suivante :

$$\hat{f} = \min_F \sum_i (d^2(m_{li}, Fm_{ri}) + d^2(m_{ri}, F^t m_{li})), \quad (2.29)$$

qui opère une minimisation simultanée sur les deux images.

Soit $l_{li} = Fm_{ri} = [l_{l1}, l_{l2}, l_{l3}]^t$ et $l_{ri} = F^t m_{li} = [l_{r1}, l_{r2}, l_{r3}]^t$. En utilisant l'équation 2.28 et $m_i^t F m_r = m_r^t F^t m_l$, la relation 2.29 s'écrit

$$\hat{f} = \min_F \sum_i w_i^2 (m_{li}^t F m_{ri})^2, \quad (2.30)$$

avec

$$w_i = \left(\frac{1}{l_{l1}^2 + l_{l2}^2} + \frac{1}{l_{r1}^2 + l_{r2}^2} \right)^{\frac{1}{2}}.$$

Méthode itérative

L'estimation de 2.30 peut être faite par une technique de moindres carrés pondérés. Ainsi la multiplication du critère linéaire par le coefficient de pondération w_i revient à remplacer les coefficients u_i dans l'équation 2.19 par $w_i u_i$, puis résoudre 2.21 par une méthode de moindres carrés linéaires. Comme les coefficients w_i dépendent de la matrice fondamentale, une solution initiale pour F est nécessaire pour commencer le processus itératif, cette solution étant obtenue en choisissant au départ les $w_i = 1$. Toutefois cette méthode donne des résultats équivalents à la méthode linéaire classique. La raison principale est que la contrainte de rang 2 de la matrice fondamentale n'est pas prise en compte.

Paramétrage non linéaire

Pour tenir compte du rang de la matrice fondamentale, il a été proposé (Luong et Faugeras, 1996; Zhang, 1996) d'utiliser des paramétrages particuliers de celle-ci. L'objectif est de choisir la matrice 3×3 , de rang 2, qui minimise 2.30. Une technique consiste à exprimer certains paramètres de la matrice fondamentale en fonction des autres. Le critère à minimiser en 2.30 devient non linéaire. Une solution initiale est nécessaire pour résoudre cette minimisation non linéaire. Elle est obtenue par l'une des méthodes linéaires décrites plus haut.

2.4.2.1 Méthode du gradient

Les techniques de moindres carrés donnent de bons résultats lorsque la variance pour chaque terme est presque la même. Comme chaque $f_i = m_{ii}^t F m_{ri}$ a une variance $\sigma_{f_i}^2$ différente, nous cherchons donc à minimiser :

$$\min_F \sum_i f_i^2 / \sigma_{f_i}^2. \quad (2.31)$$

Une approximation de la variance de f_i au premier ordre donne,

$$\sigma_{f_i}^2 = \sigma^2 [l_{i1}^2 + l_{i2}^2 + l_{r1}^2 + l_{r2}^2]$$

et 2.31 devient

$$\min_F \sum_i (m_{ii}^t F m_{ri})^2 / g_i^2,$$

avec $g_i = \sqrt{l_{i1}^2 + l_{i2}^2 + l_{r1}^2 + l_{r2}^2}$ représentant le gradient de f_i (Luong et Faugeras, 1996).

2.4.2.2 Techniques robustes

Etant donné que les mauvais appariements influencent de façon très négative les résultats du calcul de la matrice fondamentale, plusieurs auteurs (Zhang, 1996; Luong et Faugeras, 1996; Torr et Murray, 1997; Torr et Zissermann, 1997; Hartley, 1997c) se sont penchés sur l'utilisation des techniques de statistiques robustes appelées *méthodes*

de régression robustes (Rousseeuw et Leroy, 1987). Ces méthodes ont été utilisées dans le passé en vision par ordinateur dans la segmentation du mouvement (Mitiche, 1994), la reconstruction à partir de données obtenues par les senseurs (Zhang et al., 1996) et plus récemment en l'estimation de la géométrie épipolaire (Luong et Faugeras, 1996). L'attrait de ces techniques, est qu'elles permettent d'éliminer les fausses mesures si leur nombre est faible. L'objectif est de retrouver les appariements qui dévient le plus du modèle épipolaire, et de les éliminer avant d'utiliser une procédure de minimisation sur l'un des critères cités plus haut. Deux des méthodes robustes les plus utilisées sont présentées par la suite.

M-estimateurs

Les techniques standards de moindres carrés essaient de minimiser la différence entre la valeur d'une observation et sa valeur estimée par un modèle donné, c'est-à-dire minimiser le carré des résidus r_i pour chaque observation i , $\min \sum_i r_i^2$. Cette technique est instable lorsqu'il y a de fausses mesures dans l'ensemble des données. Les M-estimateurs permettent de réduire les effets des fausses mesures en remplaçant le carré des résidus par une autre fonction des résidus,

$$\min \sum_i \rho(r_i).$$

Cette fonction peut être écrite sous la forme d'une minimisation pondérée,

$$\min \sum_i w(r_i^{(k-1)}) r_i^2,$$

où $w(x) = \psi(x)/x$ représente le coefficient de pondération, et $\psi(x) = d\rho(x)/dx$ est appelée fonction d'influence. L'indice k désigne l'itération courante. Le choix de la fonction ρ est très important. Elle doit être symétrique, définie positive avec un minimum unique en zéro, et elle doit diverger moins vite que la fonction carré. Plusieurs choix de fonction pour ρ sont présentés dans la littérature, la plus utilisée est la fonction de Tukey :

$$r_i = \begin{cases} (c^2/6) \left(1 - [1 - (r_i/c\sigma)^2]^3\right) & \text{si } |r_i| \leq c\sigma \\ (c^2/6) & \text{sinon} \end{cases}$$

où σ représente la déviation standard du bruit et c est une constante.

Moindre médiane des carrés

La méthode la moindre médiane des carrés est la méthode la plus utilisée en détection et élimination des fausses mesures dans plusieurs techniques de vision. Ici elle est utilisée pour estimer les paramètres de la matrice fondamentale en résolvant le problème d'optimisation non linéaire suivant :

$$\hat{f} = \min \text{median}_i r_i^2.$$

Cet estimateur permet d'obtenir la plus petite valeur pour la médiane des carrés des résidus sur l'ensemble des données. Pour cela il faut chercher dans l'espace des points ; ceux qui donnent la meilleure estimation possible. Comme l'espace est très large, une méthode basée sur une recherche dans l'espace de manière aléatoire est utilisée. La procédure pour calculer cette estimation est décrite dans l'algorithme suivant :

Pour un appariement de n points donnés: $(m_{li}, m_{ri}) \quad i = 1, \dots, n$.

1. Une technique de Monte-Carlo est utilisée pour choisir un nombre m de sous-ensembles formés chacun par 7 points.
2. Pour chaque sous-ensemble J , on calcule la matrice fondamentale F_J .
3. Pour chaque F_J , on calcule la médiane des résidus carrés, M_J , par rapport à l'ensemble des points correspondants,

$$M_J = \text{median}_{i=1 \dots n} [d^2(m_{li}, F_J m_{ri}) + d^2(m_{ri}, F_J^t m_{li})].$$

4. Choisir F_J pour laquelle M_J est minimale.

Remarque

Dans la section précédente nous avons revu les principales méthodes de calcul de la matrice fondamentale utilisées dans différents domaines reliés à la vision par ordinateur. Récemment, les chercheurs se sont de plus en plus intéressés à l'évaluation des performances des techniques de calcul de la matrice fondamentale existantes (Torr et

Murray, 1997; Zhang, 1996; Csurka et al., 1997). Les différentes études montrent que ces techniques sont souvent très instables : l'existence de faux appariements influence d'une manière significative les résultats de ces algorithmes. Ainsi, on relève que les méthodes linéaires sont les moins performantes relativement aux critères de précision, de stabilité et de robustesse aux erreurs induites par les faux appariements. Par contre, les méthodes non linéaire donnent de meilleurs résultats. Toutefois, ils sont aussi très instables par rapport aux erreurs d'appariement. Les méthodes utilisant les statistiques robustes sont les plus performantes mais au détriment du temps machine qui est souvent exorbitant. Tout cela nous a amené à développer une nouvelle technique linéaire de calcul de la matrice fondamentale qui exploite l'information existante dans la scène à savoir la présence de surfaces planes. Ces surfaces vont nous permettre d'ajouter plus de stabilité au processus de calcul tout en gardant l'élégance de l'utilisation d'une technique linéaire permettant un gain en temps de calcul appréciable. Dans la suite nous décrivons la méthode proposée.

2.5 Approche proposée : Exploitation de la contrainte de planarité

Nous avons développé une nouvelle méthode linéaire qui exploite des informations additionnelles sur l'environnement, qui pourraient être disponibles dans certains cas, pour résoudre la contrainte épipolaire et calculer la matrice fondamentale de façon

robuste (Akhroufi, Tong, Polotski et Cohen, 1998). Ici nous allons utiliser la *contrainte de planarité* reliée à la présence de surfaces planes dans la scène.

La projection d'un plan Π , de l'espace dans un plan image, est définie par une transformation projective T , appelée *colinéation*. Cette transformation projette un plan physique de la scène Π , en un plan π dans le plan image (figure 2.7).

Si on prend un point $M = [X, Y, Z]^t$ du plan Π , sa projection dans le plan image gauche est définie par,

$$m_l = TM,$$

c'est-à-dire

$$\begin{bmatrix} x_l \\ y_l \\ 1 \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}, \quad (2.32)$$

où $m_l = [x_l, y_l, 1]^t$ est la représentation canonique du vecteur de coordonnées homogènes qui résulte de la projection du point M dans le plan image gauche.

De même, si $m_r = [x_r, y_r, 1]^t$ est la représentation canonique du vecteur de coordonnées homogènes résultant de la projection de M dans le plan image droit, il existe une *colinéation* T' qui représente la transformation projective 3D-2D du plan Π dans l'image de droite,

$$m_r = T'M,$$

c'est-à-dire

$$\begin{bmatrix} x_l \\ y_l \\ 1 \end{bmatrix} = \begin{bmatrix} t'_{11} & t'_{12} & t'_{13} \\ t'_{21} & t'_{22} & t'_{23} \\ t'_{31} & t'_{32} & t'_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}. \quad (2.33)$$

En combinant 2.32 et 2.33, on obtient la relation qui lie les deux points correspondants m_l et m_r (figure 2.7) :

$$\begin{aligned} m_r &= T'T^{-1}m_l \\ m_r &= Hm_l, \end{aligned} \quad (2.34)$$

où $H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$ représente une transformation homogène qui applique

l'image d'un plan, se trouvant sur l'image de gauche notée π_l , en l'image d'un autre plan notée π_r et qui se trouve sur le plan image droit. π_l et π_r sont les images d'un même plan physique Π , de la scène. H est appelée *homographie*. Nous avons vu dans la section 2.1.7 que H est déterminée à un facteur multiplicatif près. Nous utiliserons l'homographie entre plans présents dans les deux vues pour calculer la matrice fondamentale. Dans ce qui suit, nous allons décrire l'algorithme utilisé.

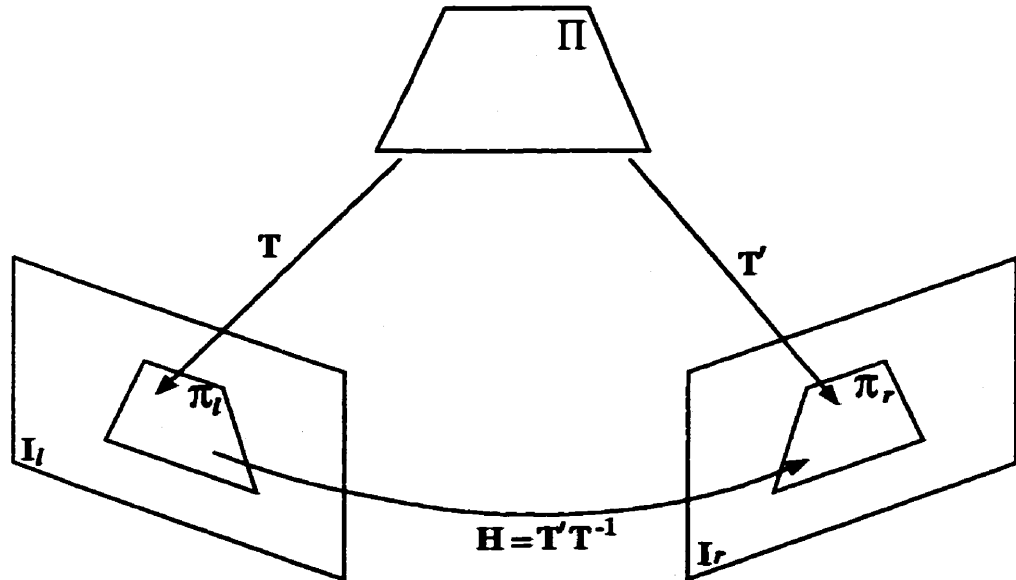


Figure 2.7 : Homographie entre surfaces planes

2.5.1 Description de la méthode proposée

Soient deux images obtenues par une paire stéréo (image droite et image gauche). Si on prend deux plans dans la scène, Π_1 et Π_2 , leurs images pour la caméra gauche sont notées π_{l1} et π_{l2} respectivement, et leurs images pour la caméra droite sont notées π_{r1} et π_{r2} . En utilisant la relation 2.34, on peut établir la relation entre π_{l1} et π_{r1} et entre π_{l2} et π_{r2} en résolvant les équations suivantes :

- Pour le plan Π_1 :

$$m_{ri} = H_1 m_{li}, \quad (2.35)$$

où m_{ri} et m_{li} sont les coordonnées des points images correspondants aux points M_i du plan Π_1 . H_1 est calculée à partir de la correspondance entre les projections d'au moins 4 points du plan Π_1 , en utilisant l'une des techniques décrites dans la section

2.1.7.

- Pour le plan Π_2 :

$$\widetilde{m}_{ri} = H_2 \widetilde{m}_{li}, \quad (2.36)$$

où \widetilde{m}_{ri} et \widetilde{m}_{li} les coordonnées des points images correspondants aux points \widetilde{M}_i du plan Π_2 . H_2 est estimée à partir d'au moins 4 points appariés, appartenant au même plan Π_2 .

La ligne épipolaire

Pour tout point $m_l(i, j)$, $i = 1, \dots, N$: et : $j = 1, \dots, M$ (où (N, M) est la taille de l'image) dans le plan image gauche, on peut calculer la ligne épipolaire qui lui correspond dans l'image droite, en utilisant les homographies. Si on considère un rayon qui passe par le centre optique de la caméra gauche et par le point image $m_l(i, j)$, il doit rencontrer le plan Π_1 en P_1 et le plan Π_2 en P_2 (figure 2.8). Puisque les points images de P_1 et P_2 coïncident en un même point image $m_l(i, j)$ dans le plan image gauche, leurs coordonnées homogènes sont les mêmes. Leurs points images dans le plan image droit sont obtenus de la manière suivante :

- De l'équation 2.35 on a :

$$m_r(1) = k_1 H_1 m_l.$$

- De l'équation 2.36 on a :

$$m_r(2) = k_2 H_2 m_l.$$

La ligne passant par $m_r(1)$ et $m_r(2)$ dans le plan image droit est la ligne épipolaire correspondant à $m_l(i, j)$. Son équation est donnée par

$$l_r = m_r(1) \otimes m_r(2),$$

où \otimes désigne le produit vectoriel.

Le point épipolaire

Pour chaque point $m_l(i, j)$ dans le plan image gauche, on a la ligne épipolaire $l_r(i, j)$ qui lui correspond dans le plan image droit. Ces lignes épipolaires s'intersectent en un point du plan image droit, appelé point épipolaire ep^r (figure 2.8). Ses coordonnées homogènes sont obtenues pour chaque paire de lignes par la relation suivante :

$$ep^r = l_r(i, j) \otimes l_r(k, l) \quad i, k = 1, \dots, N, \quad j, l = 1, \dots, M; \quad i \neq k, j \neq l.$$

En prenant la moyenne des valeurs obtenues pour l'ensemble $\{ep^r\}$, on détermine le point épipolaire ep^r . Lorsque les lignes épipolaires sont parallèles, le point épipolaire se trouve à l'infini dans le plan image droit. Néanmoins, on peut toujours le représenter en utilisant les coordonnées homogènes.

De la même façon on peut trouver le point épipolaire ep^l sur le plan image gauche.

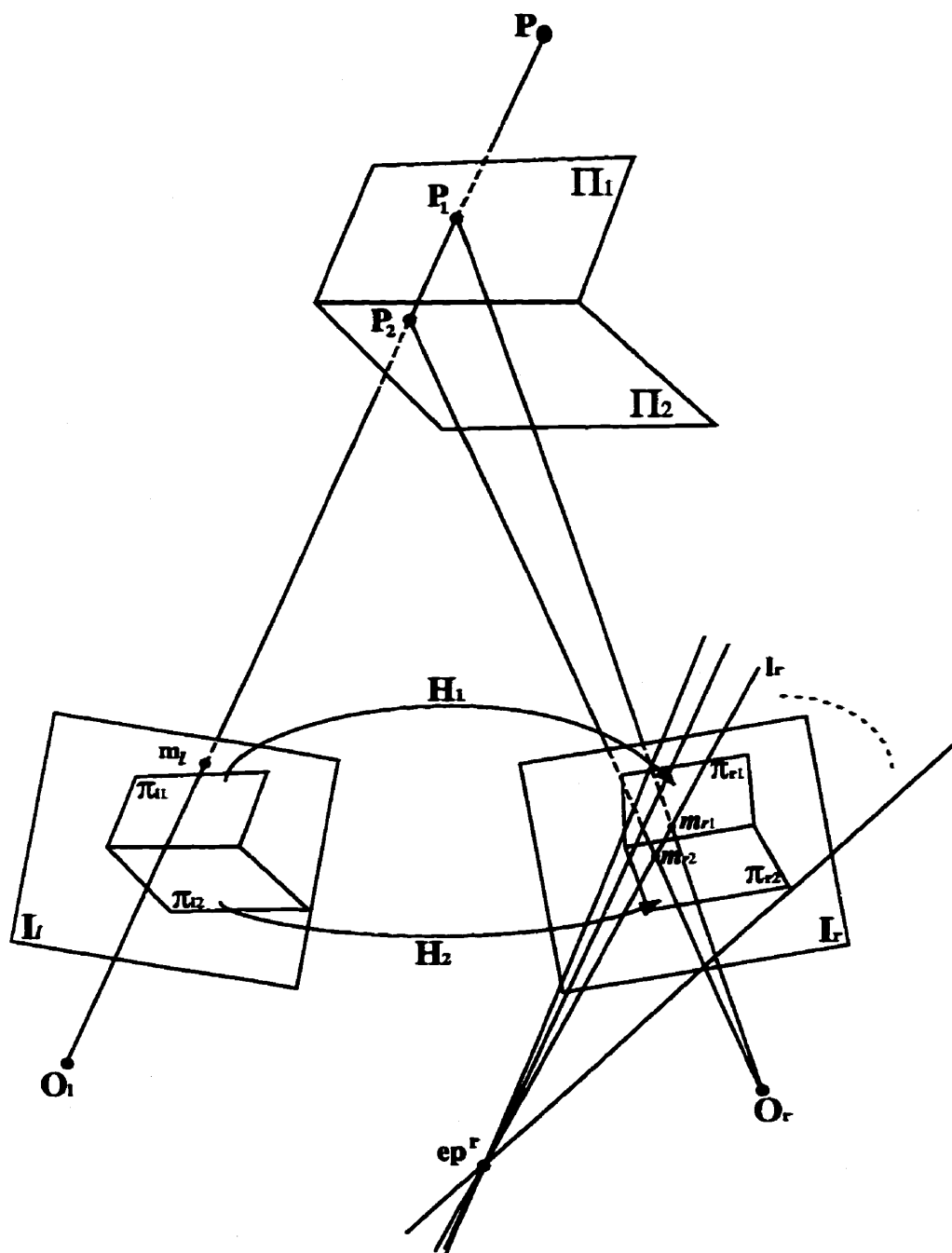


Figure 2.8 : Contrainte de planarité et géométrie épipolaire

2.5.2 Relation avec la matrice fondamentale

Pour un point m_l sur le plan image gauche, la matrice fondamentale F définit la ligne épipolaire l_r sur le plan image droit qui lui correspond. Nous avons la relation suivante (voir section 2.4.1.6) :

$$F^t m_l = l_r. \quad (2.37)$$

Posons $F = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}$, où les f_i représentent les lignes de la matrice F , et $l_r = \begin{bmatrix} l_1 & l_2 & l_3 \end{bmatrix}^t$, où les l_i représentent les coordonnées homogènes de la ligne l_r .

Soit le point épipolaire sur le plan image droit défini par le vecteur de coordonnées homogènes :

$$ep^l = [ep^l(1), ep^l(2), ep^l(3)]^t.$$

Avec les contraintes épipolaires on peut expliciter la contrainte de rang 2 pour la matrice fondamentale. On distingue deux cas particuliers : lorsque les lignes épipolaires se croisent en un point à distance finie, $ep^l(3) \neq 0$, et lorsqu'elles sont parallèles, c'est-à-dire qu'elles se croisent à l'infini, $ep^l(3) = 0$.

2.5.2.1 Cas des épiholes à distance finie

Ici nous considérons le cas où toutes les lignes épipolaires s'intersectent en un point à distance finie sur le plan image.

Proposition 1

Soit $ep^l(3) \neq 0$. Posons :

$$ep_x^l = \frac{ep^l(1)}{ep^l(3)}, \quad ep_y^l = \frac{ep^l(2)}{ep^l(3)}.$$

(ep_x^l, ep_y^l) sont les coordonnées cartésiennes du point épipolaire dans le plan image gauche. Nous avons alors la relation suivante :

$$f_3 = -ep_x^l f_1 - ep_y^l f_2. \quad (2.38)$$

Cette expression nous permet de définir la contrainte de rang 2 en fonction des coordonnées du point épipolaire.

Preuve

Nous allons montrer ici comment obtenir la relation 2.38 à partir des relations définies par les contraintes de la géométrie épipolaire. Nous avons la relation suivante entre la matrice fondamentale et le point épipolaire :

$$Fep^l = 0, \quad (2.39)$$

où ep^l est le point épipolaire dans le plan image gauche. Le vecteur de coordonnées du point épipolaire est orthogonal aux lignes de la matrice fondamentale : $ep^l \perp f_1$,

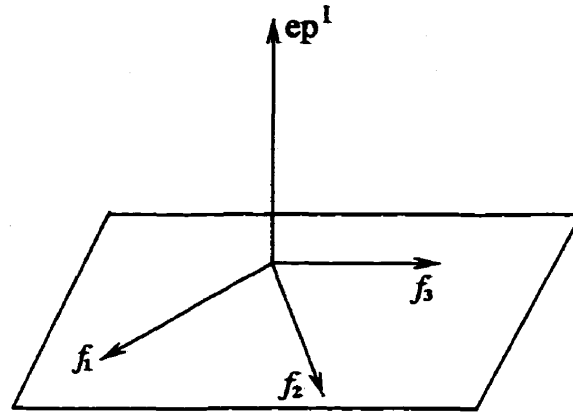


Figure 2.9 : Contrainte définie par la matrice fondamentale et le point épipolaire $ep^l \perp f_2$ et $ep^l \perp f_3$. f_1 , f_2 et f_3 sont sur un même plan perpendiculaire au vecteur représenté par le point épipolaire ep^l (figure 2.9).

Choisissons une représentation normalisée pour ep^l , f_1 , f_2 et f_3 . Nous avons :

$$ep^l = k \overline{ep^l},$$

$$f_1 = k_1 \overline{f_1},$$

$$f_2 = k_2 \overline{f_2},$$

où k , k_1 et k_2 sont des constantes. Prenons le cas où $ep^l \neq 0$, $f_1 \neq 0$ et $f_2 \neq 0$. Soit une matrice A telle que :

$$A \overline{ep^l} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad A \overline{f_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad A \overline{f_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

et

$$Af_3 = af_1 + bf_2. \quad (2.40)$$

En multipliant l'équation 2.37 par A nous obtenons :

$$AF^t m_l = Akl_r,$$

c'est-à-dire

$$A \begin{bmatrix} f_1^t & f_2^t & f_3^t \end{bmatrix} m_l = Akl_r. \quad (2.41)$$

A partir de la relation 2.40 nous obtenons

$$Af_3 = af_1 + bf_2 = \begin{bmatrix} ak_1 \\ bk_2 \\ 0 \end{bmatrix}.$$

L'équation 2.41 s'écrit

$$\begin{bmatrix} k_1 & 0 & ak_1 \\ 0 & k_2 & bk_2 \\ 0 & 0 & 0 \end{bmatrix} m_l = Akl_r.$$

Posons

$$A = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}.$$

Nous obtenons alors

$$\frac{a_1 l_r}{a_2 l_r} = \frac{k_1 m(1) + a k_1 m(3)}{k_2 m(2) + b k_2 m(3)}.$$

Soit

$$x = \frac{m(1)}{m(3)}, y = \frac{m(2)}{m(3)},$$

où (x, y) sont les coordonnées cartésiennes du point image m_i dans le plan image gauche. Nous avons donc :

$$\frac{a_1 l_r}{a_2 l_r} = \frac{k_1 x + a k_1}{k_2 y + b k_2} = \frac{k_1 x + a}{k_2 y + b}. \quad (2.42)$$

Soit un point m_0 de l'image gauche. Tous les points se trouvant sur la ligne épipolaire l_i qui passe par m_0 et ep_i , se trouvent sur la même ligne épipolaire dans le plan image droit. Pour m_0 se trouvant sur la ligne épipolaire l_i , nous avons alors :

$$\frac{a_1 l_r}{a_2 l_r} = \frac{k_1 x_0 + a}{k_2 y_0 + b}, \quad (2.43)$$

avec

$$x_0 = \frac{m_0(1)}{m_0(3)}, y_0 = \frac{m_0(2)}{m_0(3)}.$$

A partir des deux équations 2.42 et 2.43 nous obtenons

$$\frac{x_0 + a}{y_0 + b} = \frac{x + a}{y + b}.$$

C'est l'équation de la ligne épipolaire l_i (qui passe par m_0 et ep^l). Soit un point m_1 qui n'appartient pas à la ligne épipolaire l_i passant par m_0 et ep^l , nous avons alors l'équation de la ligne épipolaire qui passe par le point image m_1 et ep^l :

$$\frac{x_1 + a}{y_1 + b} = \frac{x + a}{y + b}.$$

Ces deux lignes épipolaires s'intersectent en un point que nous noterons $(-a, -b)$.

Ce point correspond au point épipolaire ep^l dans le plan image gauche :

$$a = -ep_x^l, \quad b = -ep_y^l.$$

De l'équation 2.40, nous obtenons alors la relation définie dans la proposition 1 :

$$f_3 = -ep_x^l f_1 - ep_y^l f_2,$$

ce qui complète notre preuve.

2.5.2.2 Calcul de la matrice fondamentale

En remplaçant f_3 dans l'équation 2.37 par son expression dans l'équation 2.38, nous obtenons :

$$\begin{bmatrix} f_1^t & f_2^t & -ep_x^l f_1^t - ep_y^l f_2^t \end{bmatrix} m_i = l_r. \quad (2.44)$$

Dans l'équation 2.44 les inconnues sont $f_1 = (f_{11}, f_{12}, f_{13})$ et $f_2 = (f_{21}, f_{22}, f_{23})$. Pour chaque point m_i du plan image gauche, nous obtenons deux équations de la forme :

$$\frac{f_{11}(m_1 - ep_x^l m_3) + f_{21}(m_2 - ep_y^l m_3)}{f_{13}(m_1 - ep_x^l m_3) + f_{23}(m_2 - ep_y^l m_3)} = \frac{l_1}{l_3}, \quad (2.45)$$

$$\frac{f_{12}(m_1 - ep_x^l m_3) + f_{22}(m_2 - ep_y^l m_3)}{f_{13}(m_1 - ep_x^l m_3) + f_{23}(m_2 - ep_y^l m_3)} = \frac{l_2}{l_3}. \quad (2.46)$$

Posons :

$$x = \frac{m_1}{m_3}, y = \frac{m_2}{m_3},$$

où x et y sont les coordonnées cartésiennes du point image m_i dans le plan image gauche. Les équations 2.45 et 2.46 deviennent:

$$\frac{f_{11}(x - ep_x^l) + f_{21}(y - ep_y^l)}{f_{13}(x - ep_x^l) + f_{23}(y - ep_y^l)} = \frac{l_1}{l_3}, \quad (2.47)$$

$$\frac{f_{12}(x - ep_x^l) + f_{22}(y - ep_y^l)}{f_{13}(x - ep_x^l) + f_{23}(y - ep_y^l)} = \frac{l_2}{l_3}. \quad (2.48)$$

Les équations 2.47 et 2.48 peuvent s'écrire sous forme d'équations linéaires en termes de f_{1i} et f_{2i} , $i = 1, 2, 3$. On peut calculer $f_1 = (f_{11}, f_{12}, f_{13})$ et $f_2 = (f_{21}, f_{22}, f_{23})$ à un facteur d'échelle près en résolvant un système d'équations linéaires. Ici trois appariements au moins sont nécessaires pour résoudre ce système d'équations linéaires. f_3 est obtenue en remplaçant f_1 et f_2 dans l'équation 2.38. Comme une des lignes de la matrice fondamentale est une combinaison linéaire des deux autres, la contrainte

de rang 2 est satisfaite. Ceci enlève l'étape d'approximation a posteriori de la matrice F , en imposant le rang 2 (voir section 2.4.1.5), qui est une source additionnelle de bruit, d'où la robustesse de la méthode proposée.

2.5.2.3 Cas des épipoles à l'infini

Ici nous considérons le cas des épipoles à l'infini. Dans ce cas, les lignes épipolaires sont toutes parallèles, c'est-à-dire qu'elles s'intersectent à l'infini. Ceci correspond à une transformation purement translationnelle le long des deux plans images.

Proposition 2

Soit $ep^l(3) = 0$. Posons

$$u = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}.$$

Nous avons les deux relations suivantes :

$$f_1 = ep^l(2)u, \quad (2.49)$$

$$f_2 = -ep^l(1)u. \quad (2.50)$$

Preuve

La preuve de la proposition 2 s'obtient de la même façon que pour la proposition

1.

2.5.2.4 Matrice fondamentale pour des épipoles à l'infini

En remplaçant dans 2.37 f_1 et f_2 par leurs expressions 2.49 et 2.50 nous obtenons,

$$\begin{bmatrix} ep^l(2)u^t & -ep^l(1)u^t & f_3^t \end{bmatrix} m_l = l_r. \quad (2.51)$$

Pour chaque point $[x, y]^t$ représentant le vecteur de coordonnées cartésiennes du point image gauche m_l , la relation 2.51 donne les deux équations suivantes,

$$\frac{u_1(ep^l(2)x - ep^l(1)y) + f_{13}}{u_3(ep^l(2)x - ep^l(1)y) + f_{33}} = \frac{l_1}{l_3}, \quad (2.52)$$

$$\frac{u_2(ep^l(2)x - ep^l(1)y) + f_{23}}{u_3(ep^l(2)x - ep^l(1)y) + f_{33}} = \frac{l_2}{l_3}. \quad (2.53)$$

Ici les inconnues sont les variables: $u = (u_1, u_2, u_3)$ et $f_3 = (f_{31}, f_{32}, f_{33})$.

Nous calculons u et f_3 en résolvant un système d'équations linéaires de la forme des équations 2.52 et 2.53. Trois appariements au moins sont nécessaires pour résoudre ce système d'équations linéaires. f_1 et f_2 sont déterminés par les équations 2.49 et 2.50. La matrice fondamentale F est déterminée à un facteur d'échelle près et respecte la contrainte de rang 2.

2.5.3 Cas d'un modèle de caméras stéréo calibrées

Si nous considérons que les paramètres internes de la caméra sont connus, on a alors la matrice essentielle à la place de la matrice fondamentale :

$$E = \Omega(t)R,$$

où R est la matrice de rotation et $\Omega(t)$ la matrice antisymétrique du vecteur translation.

Proposition 3

Posons

$$E = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}.$$

Si $ep^l(1) \neq 0, ep^l(2) \neq 0, ep^l(3) \neq 0$, alors le point épipolaire ep^l sur le plan image gauche est donné par

$$ep^l = \left(\frac{t_1}{t_3}, \frac{t_2}{t_3}, 1 \right)$$

et lorsque $t_3 \neq 0$, nous avons la propriété suivante :

$$e_3 = -\frac{t_1}{t_3}e_1 - \frac{t_2}{t_3}e_2. \quad (2.54)$$

Nous remarquerons que cette propriété est équivalente à la relation 2.38 que nous avons obtenue pour la matrice fondamentale pour imposer la contrainte de rang 2.

Preuve

Soit

$$R = \begin{bmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{bmatrix},$$

la matrice 3×3 de rotation, et

$$t = \begin{bmatrix} t_1 & t_2 & t_3 \end{bmatrix}^t,$$

le vecteur de translation. La matrice essentielle E est définie par (voir section 2.3) :

$$\begin{aligned} E &= \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{bmatrix}, \\ &= \begin{bmatrix} -t_3 r_{12} + t_2 r_{13} & -t_3 r_{22} + t_2 r_{23} & -t_3 r_{32} + t_2 r_{33} \\ t_3 r_{11} - t_1 r_{13} & t_2 r_{21} - t_1 r_{23} & t_3 r_{31} - t_1 r_{33} \\ -t_2 r_{11} + t_1 r_{12} & -t_2 r_{21} + t_1 r_{22} & -t_2 r_{31} + t_1 r_{32} \end{bmatrix}. \end{aligned}$$

Si nous considérons une transformation entre les caméras telle que la composante translationnelle $t_3 \neq 0$, nous vérifions facilement que :

$$\begin{aligned}
e_3 &= -\frac{t_1}{t_3} \begin{bmatrix} -t_3 r_{12} + t_2 r_{13} \\ -t_3 r_{22} + t_2 r_{23} \\ -t_3 r_{32} + t_2 r_{33} \end{bmatrix}^t - \frac{t_2}{t_3} \begin{bmatrix} t_3 r_{11} - t_1 r_{13} \\ t_2 r_{21} - t_1 r_{23} \\ t_3 r_{31} - t_1 r_{33} \end{bmatrix}^t, \\
&= -\frac{t_1}{t_3} e_1 - \frac{t_2}{t_3} e_2,
\end{aligned}$$

ce qui complète notre preuve.

Calcul de la matrice essentielle

Nous pouvons donc utiliser la même approche développée en section 2.5.2 pour calculer la matrice essentielle. Ici la contrainte de rang 2 sera imposée par la relation 2.54.

2.6 Conclusion

Après une brève introduction à la géométrie projective, nous avons abordé le problème de calcul de la matrice fondamentale. Une revue des principales méthodes a été présentée où nous avons vu que les comparaisons faites par les chercheurs montrent que les méthodes classiques souffrent d'une très grande sensibilité au bruit. Ceci est souvent dû à une mauvaise formulation de l'équation de contrainte épipolaire à résoudre, qui ne tient pas compte des propriétés de la matrice fondamentale. La contrainte de rang 2 représente la propriété la plus importante à satisfaire dans ce cas. Nous avons présenté une propriété importante reliant la matrice fondamentale et l'homographie entre les

images de surfaces planes présentes dans la scène. Cette propriété a été exploitée pour définir la géométrie épipolaire et calculer la matrice fondamentale. Ici l'homographie est utilisée pour définir, à partir des appariements, les lignes épipolaires sur chaque image. Ces lignes permettent de calculer le point épipolaire sur chaque vue. Nous avons montré que le point épipolaire peut être utilisé pour établir la contrainte de rang 2, en l'introduisant dans les coefficients de la matrice fondamentale, avant de résoudre le système de contraintes épipolaires. Ensuite, nous avons utilisé les lignes épipolaires estimées par l'homographie pour contraindre le processus de minimisation à converger vers une solution stable. Ceci nous a permis de poser le problème sous forme linéaire, permettant une résolution rapide et robuste. Des résultats seront présentés plus loin.

Chapitre 3

Invariants projectifs

Une des propriétés les plus importantes des transformations projectives est que certaines mesures géométriques sont invariantes par rapport à celles-ci. Ces mesures permettent de définir une propriété invariante d'une configuration géométrique présente dans la scène. Dans ce chapitre nous allons nous intéresser plus particulièrement aux invariants projectifs dans le plan projectif \mathcal{P}^2 et l'espace projectif \mathcal{P}^3 (Mundy et al., 1993; Maybank, 1995; Csurka et Faugeras, 1998; Shashua et Navab, 1996).

Nous allons tout d'abord introduire de manière générale la définition d'un invariant. Puis nous allons définir l'invariant fondamental en géométrie projective : le *birapport*. Nous donnerons les définitions de base du birapport et nous présenterons les différentes techniques pour le calculer à partir de points, de lignes et de plans. Nous montrerons qu'il est invariant par rapport à toute transformation projective. Nous définirons par la suite la notion de coordonnées projectives et leurs relations

avec le birapport et les coordonnées homogènes. Puis nous allons étendre la notion d'invariance aux géométries affines et euclidiennes.

Dans la dernière partie de ce chapitre nous définirons deux nouveaux invariants : le premier est basé sur deux homographies et un point de référence, et le deuxième sur une seule homographie et une ligne de référence.

3.1 Introduction aux invariants projectifs

3.1.1 Définition d'un invariant

Soient deux ensembles E_1 et E_2 et un ensemble T de transformations de E_1 dans E_2 . Soit I une fonction des éléments de E_2 . I est un invariant s'il prend la même valeur pour toutes les images d'un élément p de E_1 par n'importe quel élément de l'ensemble des transformations T :

$$\forall p \in E_1, \quad \forall t, t' \in T, \quad I(t'(p)) = I(t(p)).$$

Si les transformations sont appliquées dans un même ensemble, c'est-à-dire, $E_1 = E_2$, on a la définition suivante :

$$\forall p \in E_1, \quad \forall t \in T, \quad I(t(p)) = I(p).$$

Pour chaque point p de E on définit l'*orbite* \mathcal{O}_p comme l'ensemble des points

images de p par toutes les transformations de T ,

$$\mathcal{O}_p = \{t(p) \mid t \in T\}$$

On peut conclure que si I est un invariant pour les transformations T de E_1 dans E_2 , alors I est constant en tous les points d'une même orbite.

Un invariant est dit complet s'il est unique pour chaque orbite, c'est-à-dire si et seulement si :

$$I(f_1) = I(f_2) \iff \exists p \in E_1 \text{ tel que } f_1 \in \mathcal{O}_p \text{ et } f_2 \in \mathcal{O}_p.$$

3.1.2 Le birapport

Le birapport est l'invariant projectif fondamental à partir duquel on peut exprimer d'autres invariants. Ainsi tout invariant projectif est défini en termes de birapports. Par exemple, une conique est caractérisée par des invariants, s'exprimant en fonction du birapport, qui peuvent être calculés par de simples constructions géométriques sur les coniques.

Birapport de quatre points alignés

Soient A, B, C, D quatre points colinéaires, leur *birapport*, noté $[A, B, C, D]$, est défini comme

$$\bar{k} = [A, B, C, D] = \frac{\overline{AC} \times \overline{BD}}{\overline{AD} \times \overline{BC}}, \quad (3.1)$$

où \overline{AB} est la mesure algébrique de AB ou la distance entre les points A et B . Cette formulation du birapport peut être étendue pour les points situés à l'infini en utilisant les conventions suivantes :

$$\frac{\infty}{\infty} = 1, \quad \frac{a}{\infty} = 0, \quad \frac{\infty}{a} = \infty,$$

avec a un nombre réel quelconque.

L'invariance du birapport

Théorème

Toute transformation projective conserve le birapport.

Preuve

À travers la preuve que nous donnons du théorème, nous montrons comment on peut définir l'invariance d'un birapport par rapport à une transformation projective donnée. Pour cela on va d'abord montrer l'effet d'une transformation projective sur les coordonnées de points appartenant à une ligne. Tout point d'une ligne peut être représenté par deux coordonnées homogènes. Les points d'une ligne seront représentés par $P = [X_1, X_2]^t$, où X_1 et X_2 sont les coordonnées homogènes sur cette ligne. La

position cartésienne d'un point sur une ligne est donnée par $X = X_1/X_2$. Le modèle d'une ligne projective est donné par un ensemble de rayons passant par l'origine d'un plan. Les points sont donc l'intersection des rayons avec une ligne quelconque l dans cet espace bidimensionnel.

La transformation projective entre les lignes est donnée par une matrice, 2×2 , de transformation homogène T , telle que: $\bar{x} = TX$ où $\bar{x} = [x_1, x_2]^t$ sont les coordonnées homogènes de la ligne transformée. On a besoin de trois paramètres essentiels pour définir T à un facteur multiplicatif près. La position cartésienne d'un point sur la ligne transformée est donnée par $x = x_1/x_2$. La transformation projective d'une ligne s'écrit comme

$$x = \frac{t_{11}X + t_{12}}{t_{21}X + t_{22}}. \quad (3.2)$$

La forme de l'équation 3.2 correspond tout simplement à la représentation en coordonnées cartésiennes d'une transformation homogène dans l'espace bidimensionnel.

Soit D_{12} le déterminant de la matrice 2×2 formée par deux points, $P_1 = [X_1^1, X_2^1]^t$ et $P_2 = [X_1^2, X_2^2]^t$, se trouvant sur une même ligne l :

$$D_{12} = |P_1 P_2|,$$

qui s'écrit aussi sous l'une des formes suivantes :

$$\begin{aligned}
D_{12} &= \begin{vmatrix} X_1^1 & X_1^2 \\ X_2^1 & X_2^2 \end{vmatrix} \\
&= X_1^1 X_2^2 - X_1^2 X_2^1 = X_1^1 X_2^2 (X_1^1/X_2^1 - X_1^2/X_2^2) = X_1^1 X_2^2 (X^1 - X^2).
\end{aligned} \tag{3.3}$$

En posant $P_1 = \Lambda_1[X^1, 1]^t$ et $P_2 = \Lambda_2[X^2, 1]^t$, avec Λ_1 et Λ_2 étant les facteurs d'échelle, l'équation 3.3 devient

$$D_{12} = \Lambda_1 \Lambda_2 (X^1 - X^2). \tag{3.4}$$

Sous une transformation projective l'équation 3.4 devient $d_{12} = |T[P_1 P_2]|$, soit :

$$d_{12} = \lambda_1 \lambda_2 (x^1 - x^2) = \Lambda_1 \Lambda_2 (X^1 - X^2) |T|,$$

avec λ_1, λ_2 étant les facteurs d'échelle pour la ligne transformée. Considérons par la suite le rapport des déterminants des paires de points :

$$\bar{K} = K(P_1, P_2, P_3) = \frac{D_{31}}{D_{32}} = \frac{|P_3 P_1|}{|P_3 P_2|} = \frac{\Lambda_1 (X^3 - X^1)}{\Lambda_2 (X^3 - X^2)}.$$

Pour la ligne transformée, ce rapport devient

$$\bar{k} = k(p_1, p_2, p_3) = \frac{|p_1 p_3|}{|p_3 p_2|} = \frac{\lambda_1 (x^3 - x^1)}{\lambda_2 (x^3 - x^2)} = \frac{\Lambda_1 (X^3 - X^1)}{\Lambda_2 (X^3 - X^2)}.$$

On peut voir que le déterminant de la matrice de transformation $[T]$ a été éliminé dans le rapport. Toutefois il reste à éliminer les facteurs d'échelle; pour cela il est nécessaire d'introduire un quatrième point. Ainsi, le rapport des déterminants pour quatre paires de points permet d'obtenir un *birapport invariant* pour une transformation projective (un birapport est caractérisé par l'apparition d'un point au dénominateur et au numérateur le même nombre de fois), par exemple :

$$\begin{aligned}\bar{k} &= D_{31}D_{42}/D_{32}D_{41} = d_{31}d_{42}/d_{32}d_{41}, \\ &= \frac{(X^3-X^1)(X^4-X^2)}{(X^3-X^2)(X^4-X^1)} = \frac{(x^3-x^1)(x^4-x^2)}{(x^3-x^2)(x^4-x^1)},\end{aligned}\tag{3.5}$$

ce qui complète notre preuve du théorème.

La conservation du birapport est illustrée par la figure 3.2 , et peut s'exprimer par :

$$[A, B, C, D] = [A', B', C', D']\tag{3.6}$$

L'ordre de la ligne projective et le j-invariant

Le birapport est obtenu à partir de 4 points. Il découle de ce qui précède que la permutation de ces points donne des invariants en terme du birapport \bar{k} défini par la relation 3.5. Les quatre points donnent 4! (ou 24) différentes permutations. Ces vingt-quatre permutations donnent six invariants différents décrits par l'ensemble suivant :

$$\left\{ \bar{k}, \frac{1}{\bar{k}}, 1 - \bar{k}, \frac{1}{1 - \bar{k}}, \frac{1 - \bar{k}}{\bar{k}}, \frac{\bar{k}}{\bar{k} - 1} \right\}.$$

Comme une transformation projective générale peut amener à un changement dans l'ordre des points (figure 3.1), il est intéressant de trouver un invariant qui ne dépend pas de cet ordre. Il existe un invariant qui est indépendant de l'ordre dans lequel les points sont pris, il est appelé le *j-invariant*, et il est défini en terme du birapport \bar{k} , par

$$j(\bar{k}) = \frac{(\bar{k}^2 - \bar{k} + 1)^3}{\bar{k}^2(\bar{k} - 1)}.$$

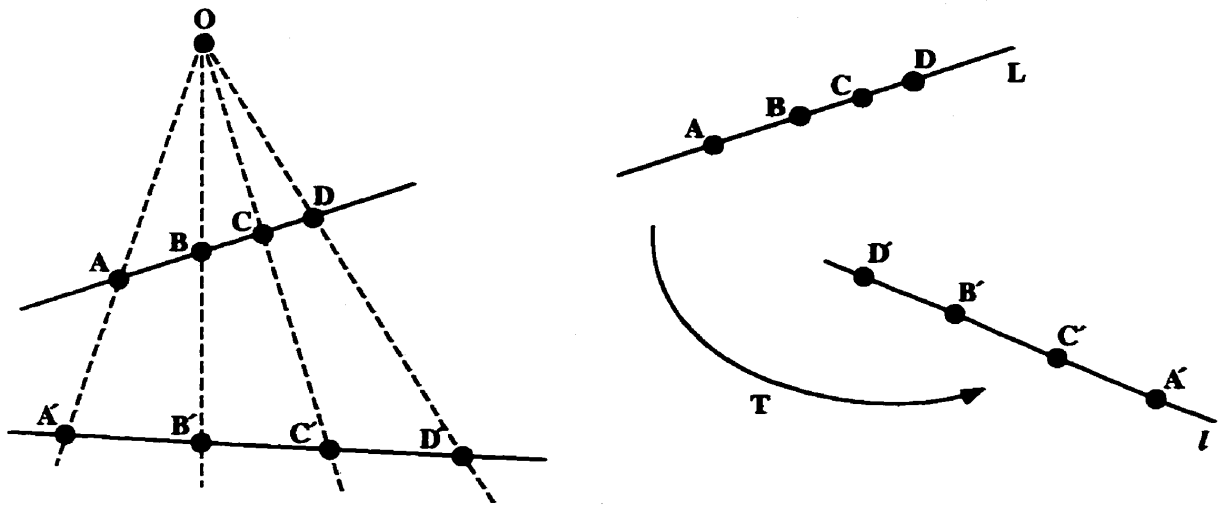


Figure 3.1 : L'ordre des points peut être changé par une transformation projective générale, il est préservé par une projection centrale

Birapport de quatre droites

Soient quatre droites l_1, l_2, l_3, l_4 concourantes en un point O et l , une ligne quelconque qui ne contient pas O et qui croise chacune des lignes du faisceau formé par les quatre droites en quatre points différents comme illustré par la figure 3.2. Le birapport du faisceau des quatre droites est défini par le birapport $[A, B, C, D]$ des points d'intersection du faisceau avec la droite l .

On le note aussi $[O; A, B, C, D]$ (birapport des droites OA, OB, OC, OD). Le birapport de quatre droites peut s'exprimer sous différentes formes. Si on utilise la forme proposée par Möbius, on peut calculer le birapport en fonction des coordonnées homogènes des points O, A, B, C, D sous la forme suivante:

$$\bar{k} = \frac{|OAC| |OBD|}{|OAD| |OBC|}, \quad (3.7)$$

où

$$|P_1 P_2 P_3| = \begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix}$$

et (x_i, y_i, z_i) sont les coordonnées homogènes de P_i .

Le birapport d'un faisceau de quatre droites peut être exprimé en fonction des angles entre les droites et peut s'écrire

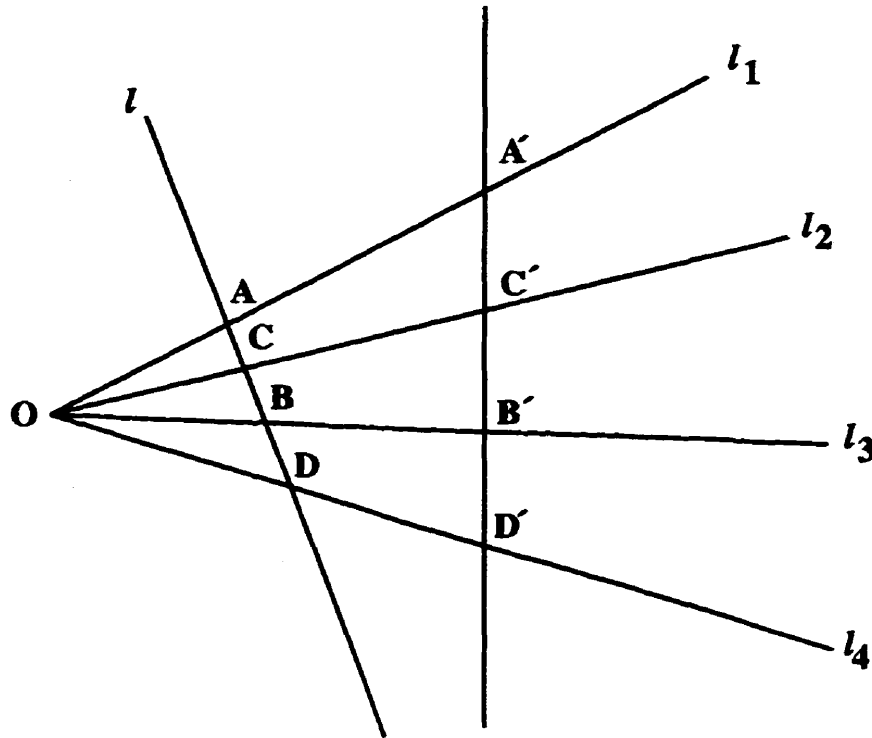


Figure 3.2 : Birapport de quatre droites

$$\bar{k} = \frac{\sin(\overrightarrow{OA}, \overrightarrow{OC'}) \sin(\overrightarrow{OB}, \overrightarrow{OD})}{\sin(\overrightarrow{OA}, \overrightarrow{OD}) \sin(\overrightarrow{OB}, \overrightarrow{OC'})}. \quad (3.8)$$

Birapport d'un faisceau de quatre plans

Pour un faisceau de quatre plans $\Pi_1, \Pi_2, \Pi_3, \Pi_4$ ayant une droite commune, le birapport est défini comme $[l_1, l_2, l_3, l_4]$, qui est le birapport de leurs quatre droites d'intersection avec un plan quelconque Π . Ceci est bien sûr indépendant du choix de Π . (voir figure 3.3).

La notion de birapport, ainsi que la formule de Möbius, peuvent être généralisées dans \mathcal{P}^n pour tout faisceau de quatre hyperplans concourants en un espace de

dimension $n - 1$.

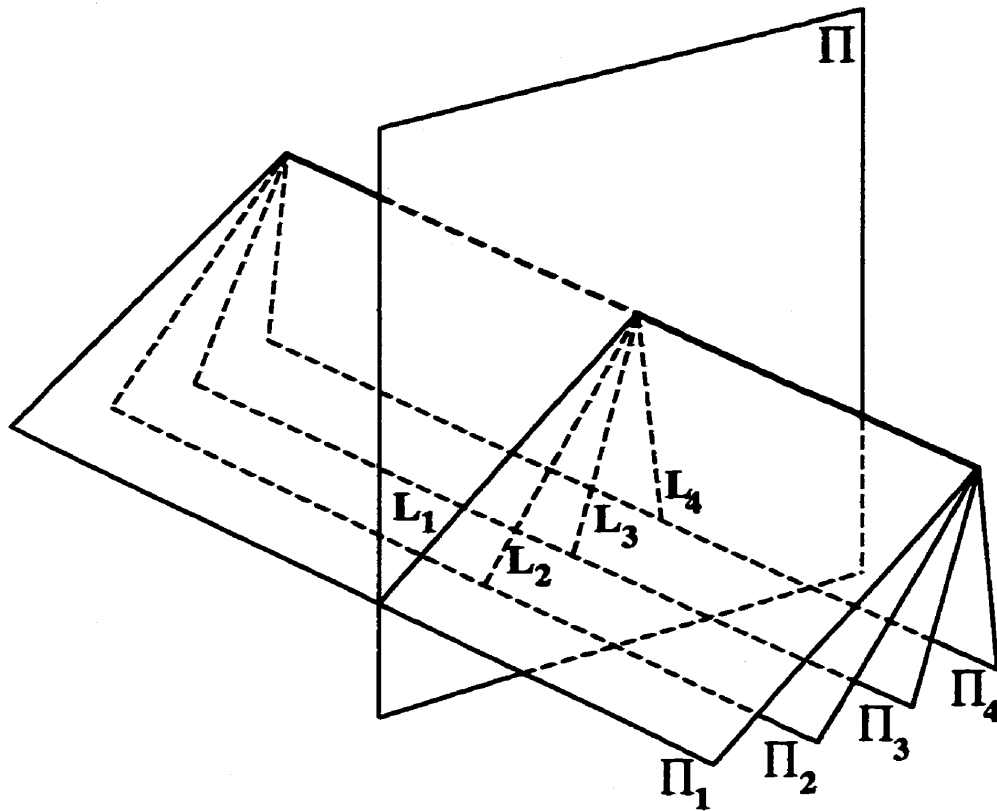


Figure 3.3 : Birapport d'un faisceau de quatre plans

3.1.3 Invariants projectifs des transformations de \mathcal{P}^3 dans \mathcal{P}^2

Les transformations projectives de \mathcal{P}^3 dans \mathcal{P}^2 sont généralement utilisées pour modéliser la projection effectuée par une caméra. Burns (Mundy et al., 1993) a démontré qu'il n'existe aucun invariant général des transformations projectives de \mathcal{P}^3 dans \mathcal{P}^2 .

Les invariants projectifs qu'on peut utiliser dans le cadre des transformations de \mathcal{P}^3 dans \mathcal{P}^2 se limitent à des ensembles de points qui ne sont pas en position générale, en particulier les ensembles de points alignés ou coplanaires. Par exemple, les n -uplets

de points coplanaires présentent des invariants pour les transformations projectives de \mathcal{P}^3 dans \mathcal{P}^2 : en effet, ils subissent une homographie de \mathcal{P}^2 . On peut donc utiliser les invariants vus précédemment, en particulier le birapport entre les plans.

3.1.4 Coordonnées projectives

La notion de coordonnées projectives est étroitement liée aux invariants projectifs. Soient P , un point de \mathcal{P}^n , et B , une base projective \mathcal{P}^n . On peut caractériser P par rapport à B par des quantités définies géométriquement à partir d'invariants projectifs : c'est ce qu'on appelle les *coordonnées projectives* de P par rapport à la base projective B . Une propriété importante de ces coordonnées est qu'elles sont invariantes par toute homographie (puisque le birapport est lui-même invariant par homographie).

Dans la suite nous donnerons une définition des coordonnées projectives dans les cas de la droite projective \mathcal{P}^1 et du plan projectif \mathcal{P}^2 , cette notion est facilement généralisable dans \mathcal{P}^n .

Coordonnées projectives sur la droite

Soient une droite l et trois points distincts (condition nécessaire pour former une base) de cette droite A , B et C formant une base projective de l . Soit P un point de l . P et A , B , C définissent un birapport, éventuellement infini. $\lambda = [A, B, C, P]$. Inversement, étant donné un scalaire λ de $\mathbb{R} \cup \{\infty\}$, il existe un unique point P tel

que le birapport $[A, B, C, P]$ soit égal à λ .

Tout couple de réels (x_1, x_2) tel que $x_1/x_2 = \lambda$ est appelé *coordonnées projectives* du point P dans la base projective définie par (A, B, C) . Le birapport permet de caractériser P . Il permet de déterminer la position d'un point P dans la base projective définie par (A, B, C) , par la relation suivante :

$$\overline{AP} = \frac{\overline{AC} \times \overline{BA}}{\lambda \overline{BC} - \overline{AC}}.$$

Lorsque $\lambda = \infty$, P est confondu avec A .

Coordonnées projectives dans le plan

Dans le plan projectif \mathcal{P}^2 , quatre points quelconques A, B, C, D non colinéaires trois à trois définissent une base projective. (voir figure 3.4).

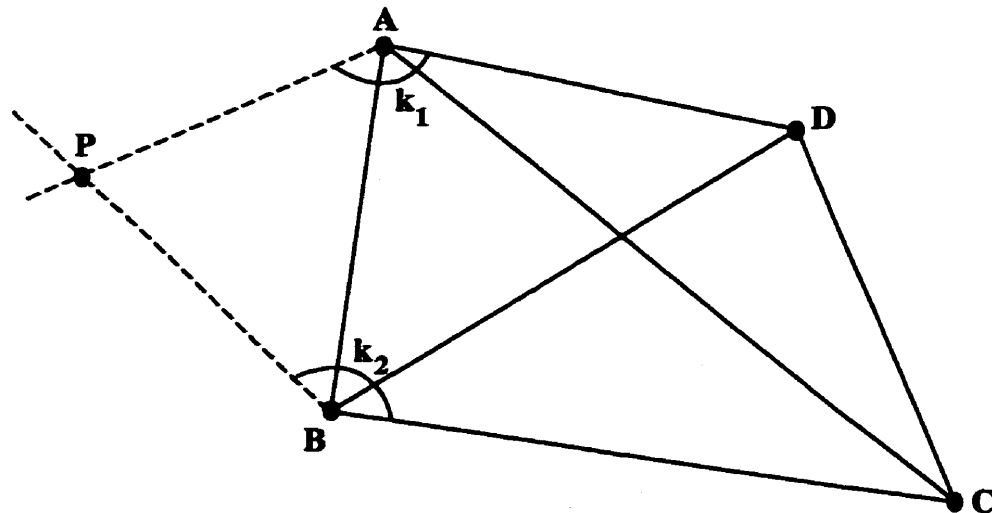


Figure 3.4 : Coordonnées projectives dans un plan

Soit un point P de \mathcal{P}^2 , alors trois nombres réels (x_1, x_2, x_3) tels que

$$x_1 = x_3[A; B, C, D, P] = x_3[AB, AC, AD, AP],$$

$$x_2 = x_3[B; A, C, D, P] = x_3[BA, BC, BD, BP],$$

sont les coordonnées projectives de P dans la base (A, B, C, D) .

Tout point P de \mathcal{P}^2 peut être défini de manière unique par ses coordonnées projectives, mis à part les points de la droite (AB) qui ont tous les mêmes coordonnées projectives. Les coordonnées projectives sont définies à un facteur multiplicatif près. Elles ne dépendent pas du système de référence des coordonnées homogènes dans lequel sont décrits les points.

Les birapports $k_1 = x_1/x_3$ et $k_2 = x_2/x_3$ suffisent pour déterminer un point quelconque en dehors de la droite (AB) .

Lien entre coordonnées projectives et coordonnées homogènes

Les coordonnées projectives se définissent par rapport à une base projective, et à un facteur multiplicatif près. Les coordonnées homogènes présentent les mêmes propriétés. Par la suite nous allons définir le lien entre ces deux systèmes de coordonnées. Etant donné un point P du plan projectif, défini par ses coordonnées projectives (x_1, x_2, x_3) dans une base projective (A, B, C, D) , soient (x, y, z) les coordonnées homogènes de P . On choisit une représentation du plan projectif où A, B, C, D ont les

coordonnées suivantes :

$$A = [0, 1, 0]^t \quad B = [1, 0, 0]^t \quad C = [0, 0, 1]^t \quad D = [1, 1, 1]^t.$$

On veut exprimer les coordonnées projectives en fonction des coordonnées homogènes. Les coordonnées (x_1, x_2, x_3) sont déterminées par la valeur des deux birapports :

$$k_1 = [A; B, C, D, P] = \frac{|ABD| |ACP|}{|ABP| |ACD|},$$

$$k_2 = [B; A, C, D, P] = \frac{|BAD| |BCP|}{|BAP| |BCD|}.$$

Comme

$$\frac{x_1}{x_3} = k_1,$$

$$\frac{x_2}{x_3} = k_2,$$

les coordonnées projectives sont donc, à un facteur multiplicatif près :

$$x_1 = \frac{|ACP|}{|ACD|} = \frac{\begin{vmatrix} 0 & 0 & x \\ 1 & 0 & y \\ 0 & 1 & z \end{vmatrix}}{\begin{vmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{vmatrix}} = x,$$

$$x_2 = \frac{|BCP|}{|BCD|} = \frac{\begin{vmatrix} 1 & 0 & x \\ 0 & 0 & y \\ 0 & 1 & z \end{vmatrix}}{\begin{vmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{vmatrix}} = y,$$

$$x_3 = \frac{|ABP|}{|ABD|} = \frac{\begin{vmatrix} 0 & 1 & x \\ 1 & 0 & y \\ 0 & 0 & z \end{vmatrix}}{\begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{vmatrix}} = z.$$

On remarque que les coordonnées projectives d'un point P dans la base projective

(A, B, C, D) sont aussi les coordonnées homogènes de ce point dans la représentation où A, B, C et D définissent les coordonnées canoniques ou : respectivement $[0, 1, 0]^t$, $[1, 0, 0]^t$, $[0, 0, 1]^t$ et $[1, 1, 1]^t$.

Les coordonnées projectives des points de la base projective ne sont pas définies. Mais on peut les définir par extension comme étant $[0, 1, 0]^t$, $[1, 0, 0]^t$, $[0, 0, 1]^t$ et $[1, 1, 1]^t$.

Ceci nous fournit une méthode pratique pour calculer les coordonnées projectives associées à un point. Etant données les coordonnées homogènes des cinq points A, B, C, D, P dans une représentation quelconque, les coordonnées projectives de P seront alors :

$$x_1 = \frac{|ACP|}{|ACD|}, \quad x_2 = \frac{|BCP|}{|BCD|}, \quad x_3 = \frac{|ABP|}{|ABD|}.$$

Inversement, on peut calculer les coordonnées homogènes d'un point en connaissant ses coordonnées projectives. En reprenant les expressions ci-dessus, on obtient trois équations linéaires en fonction des coordonnées homogènes de P . La résolution de ce système donne les coordonnées homogènes de P en fonction de ses coordonnées projectives et des coordonnées homogènes de A, B, C, D .

3.1.5 Dualité

En géométrie projective le principe de dualité nous permet d'étendre toute propriété des points en une propriété des hyperplans. Ainsi, comme nous l'avons vu au chapitre

2, les points et les droites sont duaux dans le plan projectif \mathcal{P}^2 . Cette dualité s'exprime essentiellement par l'équivalence suivante : deux points définissent une droite et deux droites définissent un point. Un ensemble de points alignés est dual d'un faisceau de droites concourantes. On sait que 5 points alignés P_1, P_2, P_3, P_4, P_5 , définissent un birapport, celui du faisceau de droites de sommet P_1 et passant par les points P_2, P_3, P_4, P_5 . Par le principe de dualité, on peut donc en déduire que 5 droites l_1, l_2, l_3, l_4, l_5 , définissent un birapport, celui des points appartenant à la droite l_1 , et qui sont les intersections de l_1 avec l_2, l_3, l_4, l_5 (voir figure 3.5).

Le principe de dualité permet de déduire directement le birapport \bar{k} , qui s'écrit :

$$\bar{k} = \frac{|l_1 l_2 l_4| |l_1 l_3 l_5|}{|l_1 l_2 l_5| |l_1 l_3 l_4|}$$

avec $|l_i l_j l_k|$ représentant le déterminant de la matrice 3×3 contenant les coordonnées homogènes des droites l_i, l_j, l_k .

3.1.6 Géométries affine et euclidienne

La géométrie projective englobe les géométries affine et euclidienne. Les transformations affines et les transformations euclidiennes sont des sous-groupes des homographies. Ainsi, les invariants projectifs sont aussi des invariants affines et euclidiens.

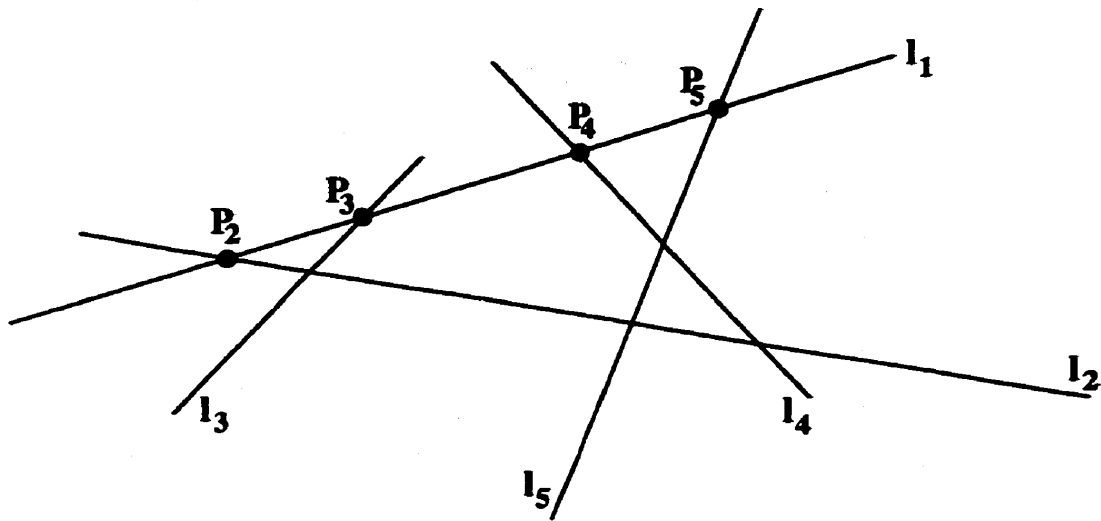


Figure 3.5 : Birapport associé à cinq droites

Géométrie affine

La géométrie affine est une restriction de la géométrie projective, elle se déduit de cette dernière en fixant l'hyperplan de l'infini.

Transformations affines

Les transformations projectives qui conservent globalement l'hyperplan de l'infini sont dites transformations affines. De ce fait, une transformation est affine si et seulement si elle conserve le parallélisme. En effet des droites sont parallèles si et seulement si elles s'intersectent en un point du plan de l'infini. Dans le cas de \mathcal{P}^n , l'hyperplan de l'infini peut donc être déterminé par n paires indépendantes de droites parallèles.

Si on fixe comme hyperplan de l'infini le plan $x_{n+1} = 0$, alors la matrice associée aux transformations affines a la forme suivante :

$$\begin{bmatrix} & t_1 \\ & \vdots \\ A & \\ & t_n \\ 0 \dots 0 & 1 \end{bmatrix}.$$

A est une matrice $n \times n$ qui correspond à la partie linéaire de la transformation affine et (t_1, \dots, t_n) est le vecteur de translation.

Invariants affines

Ici le parallélisme, qui n'est pas une notion projective (voir chapitre 2), est un invariant affine fondamental. À partir du parallélisme de droites, on peut définir la notion de plans parallèles (sécants en une droite située dans le plan de l'infini) et droites parallèles (sécantes en un point du plan de l'infini).

L'invariant algébrique fondamental en géométrie affine est le rapport des longueurs de 3 points alignés. On peut noter que cet invariant est en fait un birapport faisant intervenir un point à l'infini. Si A , B et C sont 3 points d'une droite l , et si D est le point à l'infini sur l alors, en utilisant les conventions du paragraphe 3.1.2, on a :

$$[A, B, C, D] = \frac{\overline{AC} \times \overline{BD}}{\overline{AD} \times \overline{BC}} = \frac{\overline{AC}}{\overline{BC}}.$$

Géométrie Euclidienne

La géométrie euclidienne est définie en introduisant une restriction supplémentaire de la géométrie affine : l'hyperplan de l'infini est globalement fixé, ainsi que certains points de cet hyperplan.

Transformations euclidiennes du plan

Les transformations euclidiennes du plan sont les transformations affines qui conservent une paire de points de la droite de l'infini. Ces deux points, notés I et J , sont appelés points absolus et peuvent être choisis arbitrairement sur la droite de l'infini. Une transformation euclidienne qui conserve le couple (I, J) est une transformation directe. Inversement, si le couple (I, J) a pour image le couple (J, I) , la transformation est dite indirecte.

Soient deux droites, l_1 et l_2 , et A_1, A_2 , leurs intersections respectives avec la droite de l'infini (voir figure 3.6), le birapport

$$k_{12} = [A_1, A_2, I, J]$$

est invariant par toute transformation euclidienne directe, et il est transformé en son inverse par toute transformation euclidienne indirecte.

Si on choisit $x_3 = 0$ comme droite de l'infini et comme points absolus les *points cycliques* $(1, i, 0)$ et $(1, -i, 0)$, les transformations euclidiennes sont alors des similitudes planes : rotations, translations, homothéties et symétries orthogonales. On définit

alors les cercles comme les ellipses intersectant la droite de l'infini en les points cycliques. L'angle entre deux droites l_1 et l_2 est défini par la formule de Laguerre :

$$\alpha = (\vec{l}_1, \vec{l}_2) = \frac{1}{2i} \log[A_1, A_2, I, J].$$

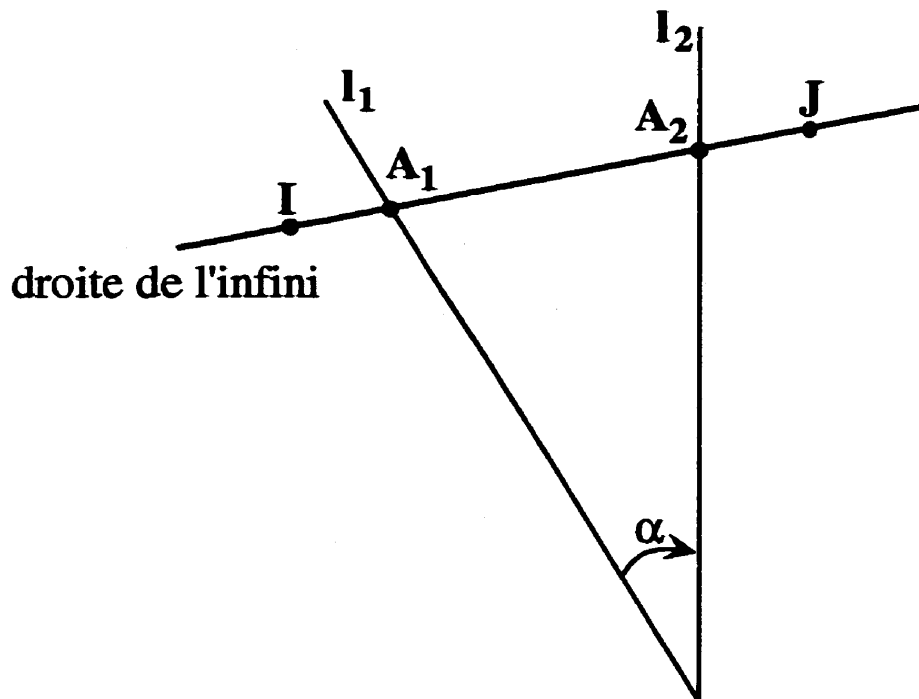


Figure 3.6 : Birapport en géométrie euclidienne

Transformations euclidiennes de l'espace

La géométrie euclidienne de l'espace est obtenue à partir de la géométrie affine de l'espace en fixant une conique du plan de l'infini. Cette conique ne contient que des points imaginaires et elle est appelée *conique absolue*; on la note Ω . Ainsi, les transformations euclidiennes de l'espace sont les transformations affines qui laissent

globalement invariante la conique absolue. Les intersections d'un plan avec Ω définissent les points absolus associés à ce plan.

Si on choisit comme plan de l'infini le plan $x_4 = 0$ et si on choisit dans ce plan la conique d'équation $x_1^2 + x_2^2 + x_3^2 = 0$ comme conique absolue, alors on retrouve ainsi, en géométrie euclidienne du plan, la notion d'angle. L'angle entre deux droites sécantes l_1 et l_2 est défini par

$$\alpha = (\vec{l}_1, \vec{l}_2) = \left| \frac{1}{2i} \log[A_1, A_2, I, J] \right|,$$

où I et J représentent les intersections du plan contenant l_1 et l_2 avec Ω (points cycliques associés à ce plan), et A_1 et A_2 sont les intersections de l_1 et l_2 avec le plan de l'infini. Ces quatre points sont alignés car ils appartiennent à la droite l_{12} , qui est l'intersection du plan contenant l_1 et l_2 et du plan de l'infini. Ils définissent un birapport (voir figure 3.7).

La conique absolue est un objet important en vision. Connaître la projection de la conique absolue sur le plan image équivaut à l'étalonnage des paramètres intrinsèques de la caméra.

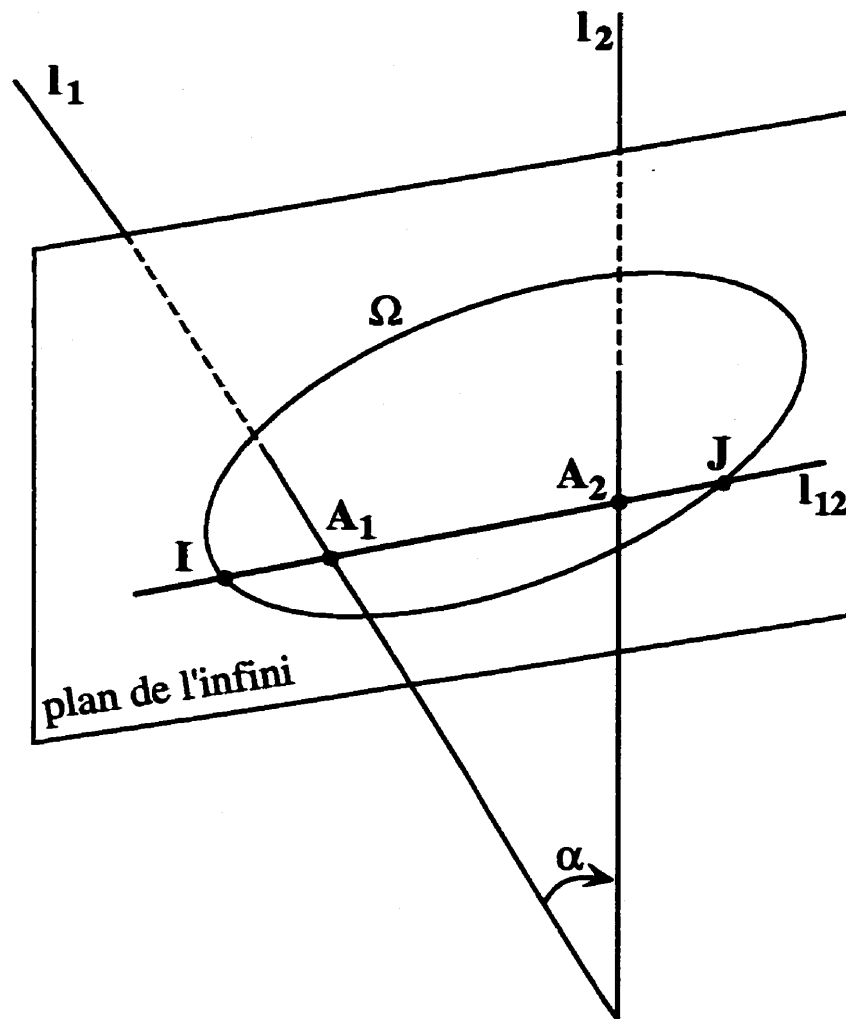


Figure 3.7 : Transformations euclidiennes de l'espace et la conique absolue

3.2 Invariant basé sur un point de référence et deux homographies

Dans cette section nous allons définir un nouvel invariant projectif que nous déduirons de la contrainte de planarité, c'est-à-dire de la présence de surfaces planes dans la scène (Tong, Akhloufi, Polotski et Cohen, 1998). Nous utiliserons deux structures

planes présentes dans l'environnement pour calculer cet invariant en termes de birapport. Nous allons utiliser la relation d'*homographie* qui existe entre deux images d'un même plan.

On sait (voir chapitre 2) qu'un plan Π de l'espace tridimensionnel est relié à sa projection, π , dans un plan image par une transformation homogène T , appelée *colinéation*. Notons par T la colinéation définie par la projection du plan Π dans le plan image gauche, et par T' la colinéation définie par la projection du même plan Π dans le plan image droit. Il existe une transformation homogène unique $H = T'T^{-1}$, qui définit la transformation entre les projections du plan Π sur les deux plans images. La matrice H est appelée *homographie* et peut être déterminée à un facteur multiplicatif près en identifiant au moins 4 paires de points images appariés correspondant au plan Π , et se trouvant sur chacune des images, π_l et π_r , de celui-ci. Dans la suite nous allons décrire l'approche proposée.

Soit P_o un point de référence choisi dans l'espace 3D (la scène). Traçons les deux lignes joignant ce point de référence aux centres optiques des deux caméras, gauche et droite, de la paire stéréoscopique, $(P_o O_l)$ et $(P_o O_r)$. Ces deux droites croisent les plans images gauche et droit en deux points, p_o^l et p_o^r respectivement. Ces deux points représentent la projection du point de référence P_o sur les deux plans images.

Soient deux plans de références Π_1 et Π_2 choisis dans la scène. L'homographie H_1 entre les images du plan Π_1 sur les deux caméras est déterminée en identifiant au moins 4 paires de points appariés appartenant à ce plan. De la même façon nous

déterminons l'homographie H_2 qui caractérise la transformation entre les images du plan Π_2 .

Considérons un point P_x de l'espace tridimensionnel, et ses projections p_x^r et p_x^l sur les plans image droit et gauche. Le point P_x peut être connecté au point de référence P_o pour former une ligne, $(P_o P_x) = L_{ox}$. Cette ligne intersecte les deux plans de référence. Le long de cette ligne on identifie quatre points: le point de référence choisi P_o , le point de la scène P_x , le point P_1 qui est l'intersection de la ligne L_{ox} avec le plan Π_1 , et le point P_2 correspondant à l'intersection de la ligne L_{ox} avec le plan Π_2 (voir figure 3.8). Nous allons noter p_1^l et p_1^r les projections du point physique P_1 sur l'image de gauche, respectivement de droite. De même, les projections de P_2 sur le plan image gauche, respectivement droit, seront notées p_2^l et p_2^r . On a alors un invariant projectif formé par les quatre points appartenant à la même ligne L_{ox} , c'est le birapport de quatre points colinéaires défini par l'équation 3.1 dans la section 3.1.2,

$$\bar{k} = [P_1, P_o, P_x, P_2]. \quad (3.9)$$

Le plan défini par la ligne L_{ox} et le centre optique O_r de la caméra droite contiennent aussi la ligne formée par p_x^r et p_o^r . Cette ligne contient aussi deux points: p_1^r et p_2^r . Cette configuration permet d'utiliser le résultat du théorème fondamental sur l'invariance du birapport par rapport aux transformations projectives, comme illustré par la figure 3.2. Ce birapport peut être également défini en utilisant les points

images de la caméra de droite, nous avons alors

$$\bar{k} = [p_1^r, p_o^r, p_x^r, p_2^r].$$

Un raisonnement équivalent nous permet d'écrire ce birapport en fonction des points images de la caméra gauche. Dans ce cas, la ligne $L_{\alpha x}$ et le centre optique de la caméra gauche, O_l forment un plan qui contient les points images p_o^l , p_1^l , p_2^l et p_x^l . Le birapport peut s'écrire comme suit :

$$\bar{k} = [p_1^l, p_o^l, p_x^l, p_2^l].$$

Dans la suite nous allons décrire comment calculer ce birapport à partir des points images appariés. Les points p_o^r et p_x^r sont connus puisqu'au départ nous choisissons notre point de référence p_o et le point p_x , pour lequel on veut calculer l'invariant. Les points p_1^r et p_2^r sont des inconnus et doivent être retrouvés en utilisant les homographies H_1 et H_2 .

Considérons l'homographie H_1 caractérisant la transformation entre les deux images du plan Π_1 . La projection de la ligne $L_{\alpha x}$ sur le plan Π_1 (voir figure 3.8), $L_{\alpha x \Pi_1}$, a une image sur chaque vue de la paire stéréoscopique. Soit $l_{\pi_1}^r$ l'image de $L_{\alpha x \Pi_1}$ dans le plan image droit. La ligne $l_{\pi_1}^r$ est obtenue en utilisant l'homographie H_1 . Nous avons alors la relation suivante :

$$l_{\pi_1}^r = [H_1^{-1}]^t l_{ox}^l, \quad (3.10)$$

où $l_{ox}^l = p_o^l \otimes p_x^l$,¹ la ligne passant par les points p_o^l et p_x^l de l'image gauche, représente la projection de la ligne L_{ox} dans l'image gauche.

Donc la projection du point P_1 sur l'image droite, p_1^r doit se trouver sur la ligne $l_{\pi_1}^r$. De plus, P_1 appartient à la ligne L_{ox} , donc son point image droit p_1^r , se trouve sur la projection de cette ligne dans l'image droite. La projection de L_{ox} sur l'image droite, notée l_{ox}^r , est la ligne passant par p_x^r et p_o^r , $l_{ox}^r = p_x^r \otimes p_o^r$. La projection du point P_1 sur la vue de droite est le point d'intersection entre la ligne $l_{\pi_1}^r$ et la ligne l_{ox}^r (voir figure 3.8). Ses coordonnées homogènes sur l'image sont données par

$$p_1^r \cong (p_o^r \otimes p_x^r) \otimes (\bar{H}_1 p_o^l \otimes \bar{H}_1 p_x^l), \quad (3.11)$$

avec $\bar{H}_1 = [H_1^{-1}]^t$.

Pour déterminer la projection du point P_2 dans l'image droite, on procède de la même façon: p_2^r est l'intersection de deux lignes, $l_{ox}^r = p_x^r \otimes p_o^r$ la projection sur l'image droite de la ligne L_{ox} , et $l_{\pi_2}^r = [H_2^{-1}]^t l_{ox}^l$ l'image droite de la projection de la ligne L_{ox} sur le plan Π_2 . On obtient alors pour p_2^r l'équation de même type que 3.11 :

$$p_2^r \cong (p_o^r \otimes p_x^r) \otimes (\bar{H}_2 p_o^l \otimes \bar{H}_2 p_x^l), \quad (3.12)$$

¹Ici \otimes représente le produit vectoriel.

avec $\overline{H}_2 = [H_2^{-1}]^t$.

Nous obtenons ainsi pour chaque point un invariant unique basé sur la sélection d'un point de référence et deux surfaces planes :

$$\overline{k} = [((p_o^r \otimes p_x^r) \otimes (\overline{H}_1 p_o^l \otimes \overline{H}_1 p_x^l)), p_o^r, p_x^r, ((p_o^r \otimes p_x^r) \otimes (\overline{H}_2 p_o^l \otimes \overline{H}_2 p_x^l))].$$

Les résultats du calcul de cet invariant à partir de l'appariement de points sur une paire d'images stéréoscopique sont illustrés dans le chapitre 5.

3.3 Invariant basé sur une ligne de référence et une seule homographie

Dans cette section, nous nous intéressons à caractériser une ligne de la scène par un invariant projectif. Nous allons présenter une nouvelle approche basée sur le choix d'une ligne de référence et un plan dans la scène (Tong et al., 1998). Tout d'abord nous commencerons par étendre la notion de birapport au cas où nous disposons seulement de deux lignes et de deux points.

Soient deux lignes, l_1 et l_2 , et deux points, x_1 et x_2 , n'appartenant pas à l_1 et l_2 (voir figure 3.9). Dans ce cas, un invariant projectif peut être défini en termes de birapports. Traçons la ligne passant par x_1 et x_2 , cette ligne coupe les lignes l_1 et l_2 en deux points différents. L'invariant correspond au birapport des distances entre les

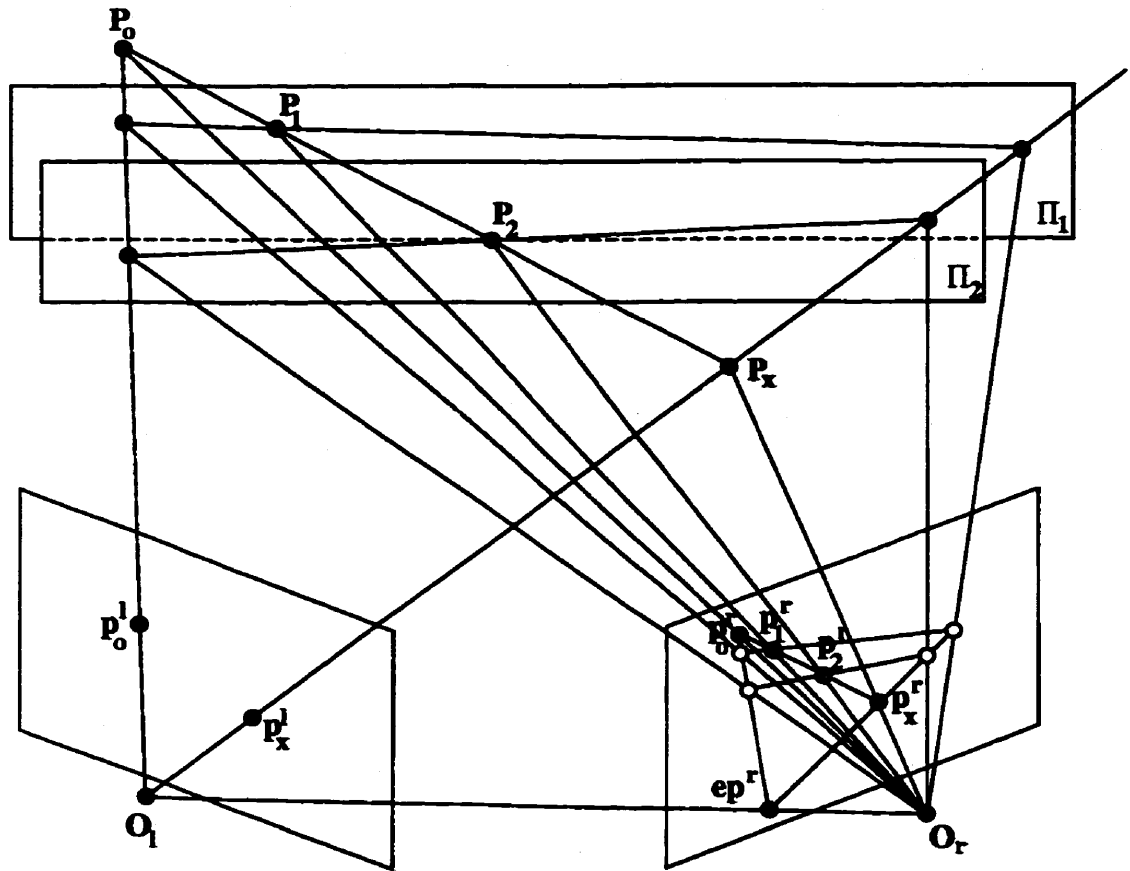


Figure 3.8 : Invariant basé sur un point et deux homographies

lignes et les points, le long de la ligne (x_1x_2) . Par exemple, la distance entre la ligne l_1 et le point x_1 est donnée par l_1x_1 . Le birapport s'écrit alors :

$$\bar{k} = \frac{(l_1x_1)(l_2x_2)}{(l_1x_2)(l_2x_1)}. \quad (3.13)$$

Nous allons utiliser la relation 3.13 pour définir un nouveau birapport basé sur une seule homographie et une ligne de référence, à partir d'une paire d'images. Nous décrivons l'approche proposée dans ce qui suit.

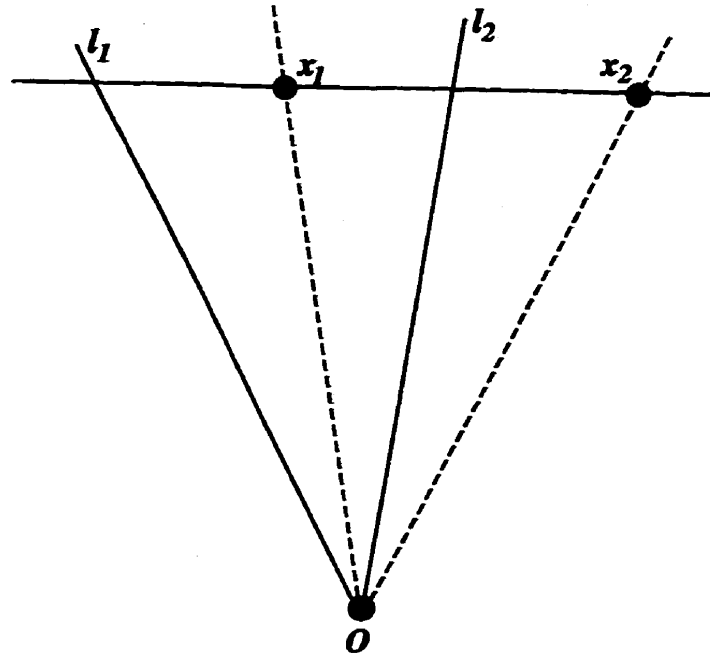


Figure 3.9 : Birapport de deux points et deux lignes

Soit une homographie H définissant la transformation homogène entre les images d'un même plan de référence, notée Π_{ref} . Soient deux lignes quelconques coplanaires, L_1 et L_2 , que nous pouvons choisir en utilisant l'homographie H . L_{ref} est une ligne de référence choisie.

Pour une ligne quelconque L_x , ses projections dans les deux plans images gauche et droit sont respectivement l_x^l et l_x^r . Les projections de la ligne de référence L_{ref} , de L_1 et de L_2 , sont respectivement $l_{ref}^l, l_{ref}^r, l_1^l, l_1^r, l_2^l$ et l_2^r . (voir figure 3.10)

La ligne de référence L_{ref} , rencontre le plan de référence Π_{ref} en un point noté P_1 . De même, l'intersection de la ligne L_x avec le plan de référence se fait en un point noté P_2 . P_1 et P_2 définissent une droite $L = (P_1P_2)$. Cette droite rencontre les lignes L_1 et L_2 en deux points différents notés respectivement X_1 et X_2 . Les points $P_1, X_1,$

X_2 et P_2 définissent un birapport. Ce birapport peut être défini en termes de deux lignes et deux points comme nous avons vu en 3.13, en utilisant les points P_1 , P_2 et les lignes L_1 et L_2 . Le birapport s'écrit sous la forme suivante :

$$\bar{k} = \frac{(L_1 P_1)(L_2 P_2)}{(L_1 P_2)(L_2 P_1)}. \quad (3.14)$$

Ce birapport peut être déduit à partir des coordonnées images des points P_1 , P_2 et lignes L_1 et L_2 . L_1 et L_2 sont choisis sur les plans images, il suffit alors de calculer les projections sur les plans images de P_1 et P_2 .

Prenons les projections sur le plan image droit par exemple. P_1 est le point d'intersection de la ligne de référence L_{ref} avec le plan de référence Π_{ref} . La projection de la ligne L_{ref} sur le plan Π_{ref} , notée ${}^{\Pi}L_{ref}$, passe aussi par le point P_1 . Ceci signifie que L_{ref} sur ${}^{\Pi}L_{ref}$ se croisent en P_1 . Donc les projections de ses deux lignes sur le plan image doivent se croiser en la projection de P_1 sur le même plan image.

La projection de L_{ref} sur le plan image droit est l_{ref}^r , et la projection de ${}^{\Pi}L_{ref}$ sur le plan image droit ${}^{\pi}l_{ref}^l$, est donnée par

$${}^{\pi}l_{ref}^r = [H^{-1}]^t l_{ref}^l,$$

avec l_{ref}^l , la projection sur le plan image gauche de la ligne de référence L_{ref} .

La projection de P_1 sur l'image de droite est alors donnée par :

$$p_1^r \cong l_{ref}^r \otimes [H^{-1}]^t l_{ref}^l.$$

De la même manière, on obtient la projection de P_2 dans l'image droite, en utilisant la projection de la ligne L_x sur le plan de référence Π_{ref} , notée ${}^\Pi L_x$. La projection de P_2 sur l'image de droite p_2^r , est donnée par l'intersection entre la projection sur le plan image droit de la ligne L_x , notée l_x^r et la projection de la ligne ${}^\Pi L_x = [H^{-1}]^t l_x^l$ sur la caméra droite :

$$p_2^r \cong l_x^r \otimes [H^{-1}]^t l_x^l,$$

où l_x^l représente la projection de la ligne L_x sur le plan image gauche.

Ainsi, le birapport donné par la relation 3.14, peut être réécrit en termes des coordonnées des points sur l'image. Pour le plan image droit, nous avons

$$\bar{k} = \frac{(l_1^r p_1^r)(l_2^r p_2^r)}{(l_1^r p_2^r)(l_2^r p_1^r)}, \quad (3.15)$$

et pour le plan image gauche, nous avons

$$\bar{k} = \frac{(l_1^l p_1^l)(l_2^l p_2^l)}{(l_1^l p_2^l)(l_2^l p_1^l)}. \quad (3.16)$$

Nous obtenons d'une manière simple un invariant projectif en terme de birapport de deux points et deux droites.

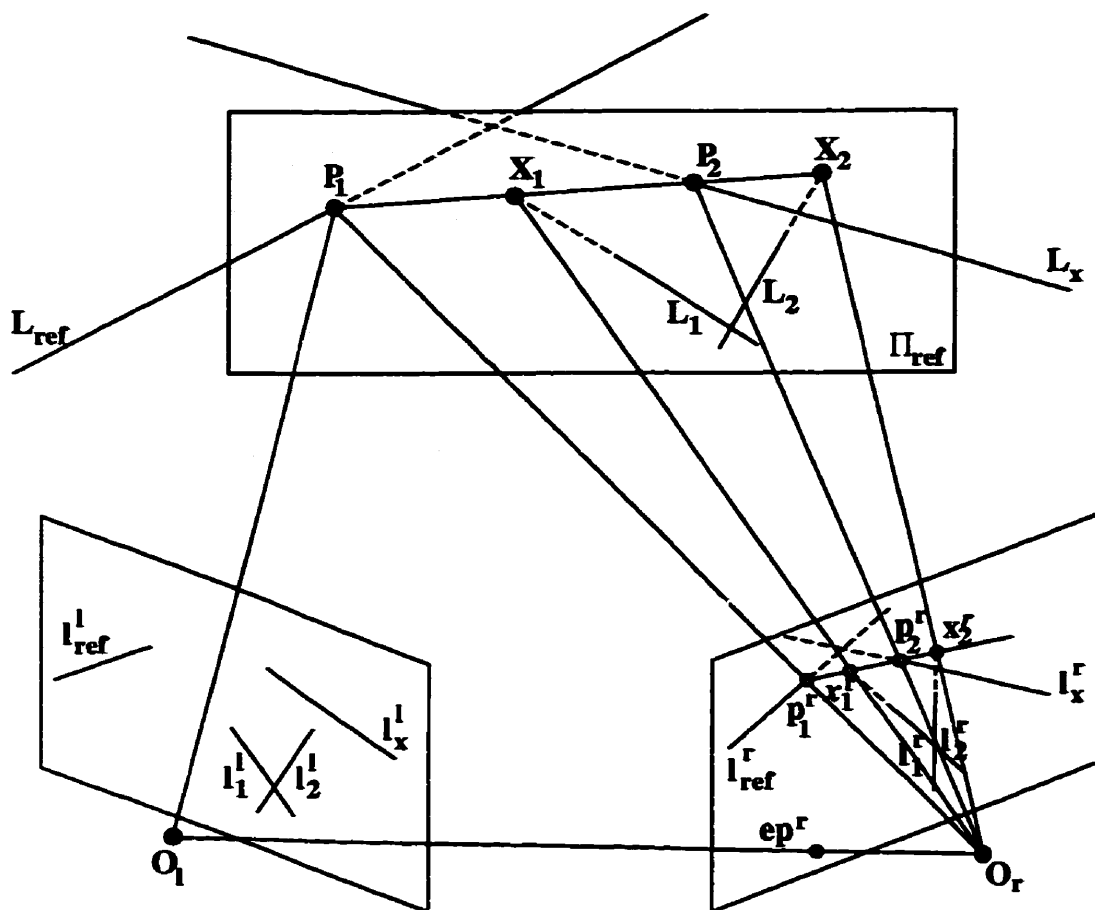


Figure 3.10 : Invariant basé sur une ligne de référence et une seule homographie

3.4 Conclusion

Dans ce chapitre nous avons introduit une théorie qui connaît une popularité croissante en vision : l'*invariance projective* (Forsyth, Mundy, Zisserman et Brown, 1990; Mundy et al., 1993; Carlsson, 1995; Hartley, 1994b; Shashua, 1993; Shashua, 1994a). Son importance vient du fait qu'elle nous procure des outils permettant de définir des invariants caractéristiques de la géométrie de la scène.

Alors que la géométrie épipolaire nous permet de caractériser les transformations

géométriques entre les images, l'invariance projective joue un rôle important dans la définition des relations géométriques mesurables qui caractérisent les propriétés d'une scène en particulier. Les invariants projectifs sont donc les invariants géométriques associés aux transformations projectives.

Après avoir défini la notion du birapport, l'invariant fondamental en géométrie projective, nous avons présenté deux nouvelles approches pour calculer l'invariant projectif dans chaque cas. Nous avons exploité l'information présente dans la scène, plus particulièrement la présence de surfaces planes, pour définir cet invariant en terme de birapports.

Les applications de l'invariant projectif peuvent être nombreuses, par exemple la reconnaissance d'objets (Mundy et al., 1993; Shashua, 1994*a*), la reconstruction tridimensionnelle de la structure de la scène (Hartley, 1994*b*), la correspondance (Shashua, 1994*a*), etc.

Dans le chapitre suivant nous allons nous intéresser plus particulièrement à l'utilisation de cet invariant pour le transfert de points sur une nouvelle vue.

Chapitre 4

Synthèse de vues

Dans ce chapitre nous allons nous intéresser à la synthèse de vues à partir d'images de référence. Plus particulièrement aux approches *géométriquement valides*.

Dans la première partie de ce chapitre nous allons réviser la bibliographie en synthèse de vues en présentant les principales techniques utilisées. Dans la deuxième partie, nous allons nous intéresser de plus près à synthétiser les vues à partir d'une paire d'images. Nous présenterons une nouvelle approche qui permet d'éviter un appariement dense entre les images lors de la synthèse de vues. Nous exploiterons les outils fournis par l'infographie pour réaliser un rendu basé sur images de bonne qualité, à partir d'un appariement épars défini par un petit nombre de primitives d'intérêt présentes dans les deux images. Dans notre approche nous tenons compte des contraintes géométriques inhérentes à la scène par l'utilisation des approches géométriques définies dans les chapitres précédents (géométrie épipolaire et invariance

projective). Nous finirons ce chapitre par une description des différentes parties de l'algorithme proposé.

4.1 État de l'art en synthèse de vues

4.1.1 Synthèse par reconstruction tridimensionnelle explicite

La reconstruction tridimensionnelle de la scène est l'un des domaines les plus actifs en vision, on trouve plusieurs articles dans la littérature traitant le sujet de la récupération de la structure de la scène à partir des images de celle-ci (Adiv, 1985; Kanade, 1995; Kanade, Yoshida, Oda, Kano et Tanaka, 1996; Kutulakos, 1995; Okutomi et Kanade, 1993; Faugeras, 1992; Huang et Netravali, 1994; Faugeras, Robert, Laveau, Csurka, Zeller, Gauclin et Zoghلامي, 1998; Bougnoux, 1998). Une fois le modèle de la scène obtenu, on peut facilement appliquer les techniques d'infographie pour produire un rendu réaliste. L'application de la texture, des effets de réflexion, de transparence, de placement de source de lumière, etc. permet d'obtenir un objet dont le réalisme approche celui du vrai objet dans la scène. La manipulation de cet objet devient très simple permettant de le visualiser sous différents angles. Plusieurs équipes de recherche se sont intéressées à cet aspect de modélisation explicite de la scène pour la synthèse de vues.

Par exemple, Koch (Koch, 1995) utilise la stéréoscopie pour la reconstruction tridimensionnelle de la scène, puis la génération de nouvelles vues. L'algorithme qu'il

propose se résume en 6 étapes :

1. étalonnage des caméras stéréoscopiques;
2. rectification de la paire d'images stéréoscopiques;
3. appariement dense des points sur les deux images;
4. triangulation des points appariés;
5. reconstruction d'un modèle en facettes planes;
6. calquage de texture sur le modèle.

Pour l'appariement dense il utilise une technique itérative qui permet un raffinement des résultats à chaque itération.

Un autre travail utilisant la modélisation 3D explicite est celui de Debevec et al. (Debevec et Malik, 1996; Taylor, Debevec et Malik, 1996a; Taylor, Debevec et Malik, 1996b). Ce travail traite exclusivement de la reconstruction de bâtiments à partir de photographies. Il propose un système de *"Conception Assistée par Ordinateur"* (CAO), qui comme la plupart des systèmes commerciaux de modélisation 3D nécessite l'intervention de l'utilisateur tout le long du processus de modélisation et de rendu. Une fois le modèle reconstruit, on lui applique la texture. Cette étape est très importante, puisque, contrairement aux systèmes de modélisation 3D classiques qui considèrent la texture comme étant statique, Debevec propose le choix de la texture parmi les images de référence les plus proches du point de vue courant. Une pondération des

textures est alors appliquée : plus l'image de référence est proche, plus le poids est grand. Un bon choix de la texture permet un gain appréciable en qualité visuelle du rendu final.

D'autres travaux, comme ceux de Niem et al. (Niem et Broszio, 1995), se basent sur la reconstruction 3D à partir des contours d'occlusion. L'objet à reconstruire est placé sur une table tournante. Niem et al. établissent un appariement dense entre les images, puis décomposent la scène en surfaces triangulaires par la triangulation des points appariés. Ces surfaces triangulaires sont ensuite texturées. Ils utilisent un calquage de textures pondéré parmi les images de référence les plus proches du point de vue à générer. Les trous dus aux occlusions sont comblés par lissage avec les voisins.

Kanade et al. (Kanade, Narayanan et Rander, 1995; Kanade, Rander et Narayanan, 1997; Narayanan et Kanade, 1998; Narayanan, Rander et Kanade, 1998; Rander, Narayanan et Kanade, 1997; Rander, Narayanan et Kanade, 1998) utilisent un système formé par une cinquantaine de caméras fixées sur un dôme, ce qui permet d'établir un appariement dense entre les images. Les caméras sont étalonnées permettant ainsi d'établir un appariement selon des directions de recherche connues. Le système est rigide, complexe et ne couvre que la scène se trouvant sous le dôme.

Il existe aussi d'autres méthodes qui évitent l'étape de la triangulation de la scène. Par exemple, Scharstein (Scharstein, 1996) fait une reconstruction dense après l'étape d'appariement dense entre les images. Une fois que les coordonnées tridimensionnelles

des points sont obtenues, il suffit de les reprojeter sur la nouvelle position du plan image.

Seitz et al. (Seitz, 1997; Seitz et Dyer, 1997*b*; Seitz et Dyer, 1997*a*; Seitz et Kutulakos, 1998) proposent l'utilisation de voxels colorés pour décrire la scène. Ces voxels sont ensuite reprojétés sur la nouvelle vue pour générer la nouvelle image.

McMillan et al. (McMillan et Bishop, 1995) utilisent un système de prises de vues qui effectue juste une rotation selon l'axe vertical passant par le centre optique de la caméra. Ainsi, les images de référence se projettent sur un repère cylindrique. La nouvelle image synthétisée est produite par la composition des images de référence utilisant une forme de reconstruction implicite des points 3D.

Plusieurs auteurs utilisent des techniques similaires pour la synthèse de vues en faisant une reconstruction tridimensionnelle dense ou partielle de la scène (Adelson, 1995; Baker, Szeliski et Anandan, 1998; Jain et Wakimoto, 1995; Moezzi, Katkere, Kurara et Jain, 1996*a*; Katkere, Moezzi, Kurara, Kelly et Jain, 1997; Moezzi, Katkere, Kurara et Jain, 1996*c*; Moezzi, Katkere, Kurara et Jain, 1996*b*; Moezzi, Li-Cheng et Gerard, 1997; Seitz et Dyer, 1995*a*; Zhang, 1998*a*).

4.1.2 Mosaïquage tridimensionnel

La représentation étendue de la scène sous forme de mosaïques constitue une approche intéressante en synthèse de vues. La réalisation de modèles étendus tridimensionnels permet d'avoir une vue plus large de la scène.

Sawhney et al. (Sawhney et Kumar, 1997) proposent d'estimer le mouvement dominant de la séquence d'images à partir des équations 4.1 liant les coordonnées des points appariés dans deux images. Les points (x_2, y_2) de l'image 2 correspondant aux points (x_1, y_1) de l'image 1 sont tels que

$$\begin{cases} x_2 = ax_1 + by_1 + c, \\ y_2 = dx_1 + ey_1 + f. \end{cases} \quad (4.1)$$

Une fois les paramètres (a, b, c, d, e, f) calculés, il est possible de procéder à un appariement dense. Les images sont ensuite reprojctées dans un même repère pour constituer une seule grande image.

Szeliski et al. (Szeliski, 1998; Kang et Szeliski, 1997; Shum, Han et Szeliski, 1998; Szeliski et Shum, 1997) utilisent une technique équivalente en commençant d'abord par estimer le mouvement dominant, puis par établir un appariement dense à base de splines.

Kumar et al. (Kumar, Anandan, Irani, Bergen et Hanna, 1995) utilisent une modélisation différente pour le mouvement dominant qui est décrit par des équations quadratiques :

$$\begin{cases} x_2 = ax_1 + by_1 + c + gx_1^2 + hx_1y_1 \\ y_2 = dx_1 + ey_1 + f + gx_1y_1 + by_1^2. \end{cases}$$

4.1.3 Rendu basé sur les images : Approches non géométriques

Les approches non géométriques sont celles où l'image synthétisée n'est pas géométriquement correcte, c'est-à-dire qu'elle ne respecte pas les contraintes imposées par la géométrie de la scène. Malgré que les images peuvent être assez réalistes, ce réalisme se limite à une zone géométrique où les contraintes ont été respectées. Plus on s'éloigne de cette zone, plus l'image se dégrade et devient inacceptable en termes de géométrie.

4.1.3.1 Interpolation de vues

Les techniques d'interpolation d'images sont très populaires en infographie, elles permettent la production d'effets spéciaux souvent spectaculaires (Heckbert, 1986; Wolberg, 1990). La technique est simple. Sur une image source on place des points de contrôle fixes. La surface entre ces points est élastique et peut être déformée par interpolation des pixels qui la forment. Ces points peuvent ensuite être déplacés pour étirer ou comprimer l'image dans certaines zones. Cela peut être appliqué à la synthèse de nouvelles vues.

A partir de deux images, source et destination, sur lesquelles on place des points de contrôle appariés, on peut générer de nouvelles vues intermédiaires en interpolant les positions de ces points de contrôle.

A partir d'un petit nombre de points de contrôle, on peut construire un mail-

lage triangulaire, puis appliquer des déformations affines des textures à l'intérieur de chaque triangle.

4.1.3.2 Combinaisons linéaires des images

Dans la littérature (Ullman et Basri, 1991; Werner, Hersch et Hlavac, 1995; Hlavac, Leonardis et Werner, 1996), on trouve une autre méthode de synthèse de vues proche des techniques d'interpolation. Il s'agit de la combinaison linéaire d'images. L'idée est que la position d'un point sur une nouvelle vue peut être estimée à partir des points qui lui correspondent dans les vues sources. La projection d'un point dans une nouvelle image est une combinaison linéaire des projections du même point dans les autres images. Cette modélisation a été proposée par Ullman et al. (Ullman et Basri, 1991), pour la reconnaissance des objets, puis utilisée par Werner et al. (Werner et al., 1995) pour la synthèse d'images. Si un point P de l'espace se projette dans n images en p_i ($i = 1 \dots n$), alors sa position dans la nouvelle vue $n + 1$, notée p_{n+1} , est donnée par l'équation suivante :

$$p_{n+1} = \sum_{i=1}^n a_i p_i.$$

Les coefficients a_i sont déterminés à partir d'un certain nombre de correspondances établies entre certains points de l'image de destination et des images sources. Dans le cas de caméras orthographiques, Werner et al. (Werner et al., 1995) proposent une relation affine, donnant la position d'un point p_3 dans une troisième vue connaissant

ses positions p_1 et p_2 dans deux vues de référence :

$$\begin{cases} p_3^x = a_1 p_1^x + a_2 p_2^x + a_3, \\ p_3^y = a_4 p_1^y + a_5 p_2^y + a_6. \end{cases}$$

4.1.4 Rendu basé sur les images: Approches géométriques

L'objectif de tout système de synthèse de vues est de produire une image qui serait semblable à celle prise réellement par une caméra si elle était placée au même endroit que celui choisi pour placer la caméra virtuelle et générer la nouvelle vue, c'est-à-dire, obtenir des vues physiquement valides. Pour cela il faut avoir un système qui tient compte des contraintes géométriques présentes dans la scène. Ici nous révisons les approches qui utilisent d'une manière ou d'une autre cette information de la scène.

4.1.4.1 Interpolation de vues valide

Seitz et al. (Seitz et Dyer, 1995b; Seitz, 1997) proposent d'utiliser l'interpolation d'images pour générer la nouvelle vue. La technique proposée respecte la contrainte d'ordre, ou monotonie. Ainsi les vues générées sont géométriquement valides. L'algorithme se divise en trois étapes (voir figure 4.1) :

1. les deux images sont rectifiées;
2. la vue intermédiaire est obtenue par interpolation linéaire entre les deux vues rectifiées;

3. la vue synthétique est dérectifiée en la reprojétant dans le plan image défini par la nouvelle position de la caméra virtuelle.

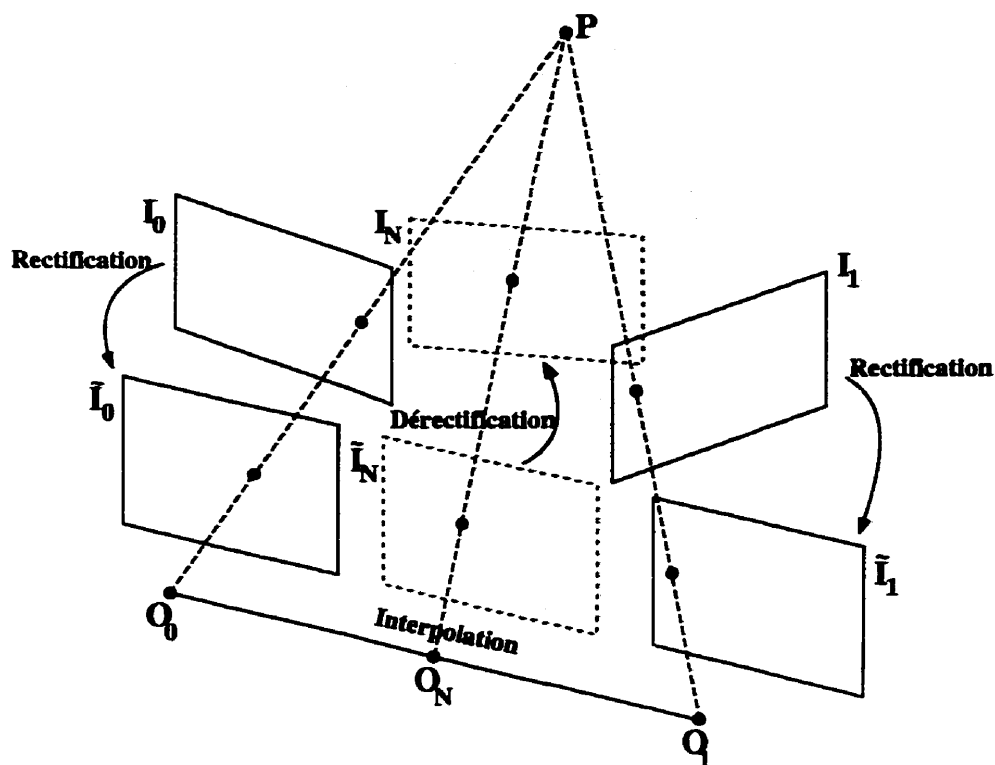


Figure 4.1 : Synthèse de vues par interpolation

La position de la caméra virtuelle est définie par la projection de quatre points de contrôle sur le plan image de la vue synthétique. Ici la caméra est étalonnée.

L'appariement dense est nécessaire pour pouvoir transférer dans l'image synthétique chaque couple apparié dans les images de référence. Ceci est réalisé de manière à respecter la contrainte de monotonie.

4.1.4.2 Utilisation de la géométrie épipolaire

Dans la littérature (Faugeras et Robert, 1996; Laveau et Faugeras, 1994; Faugeras et al., 1998), on trouve des méthodes qui utilisent explicitement la géométrie épipolaire. Le transfert se fait par intersection des lignes épipolaires. En effet, connaissant une paire de points appariés, p_1 et p_2 , correspondant à un point P de l'espace dans les images 1 et 2, et les matrices fondamentales F_{13} et F_{23} liant les images 1 et 3 (respectivement les images 2 et 3), alors le point p_3 sur l'image synthétisée 3 se trouve sur l'intersection de la ligne épipolaire correspondant à p_1 dans l'image 3 et la ligne épipolaire correspondant à p_2 dans cette même image 3; il est donné par :

$$p_3 = F_{13}p_1 \otimes F_{23}p_2. \quad (4.2)$$

La position de la caméra virtuelle est obtenue en plaçant manuellement cinq points décrivant le plan de la caméra et son centre optique. Ainsi, on évite l'étalonnage complet des caméras puisque les matrices fondamentales suffisent.

L'équation 4.2 ne peut pas être utilisée si les lignes $F_{13}p_1$ et $F_{23}p_2$ sont parallèles ou confondues. Donc, tous les points P appartenant au plan formé par les centres optiques des trois caméras, appelé plan trifocal, ne peuvent pas être reprojetés par cette technique (voir figure 4.2), et les équations deviennent numériquement instables pour des points proches du plan trifocal. L'image ne peut être synthétisée dans cette zone.

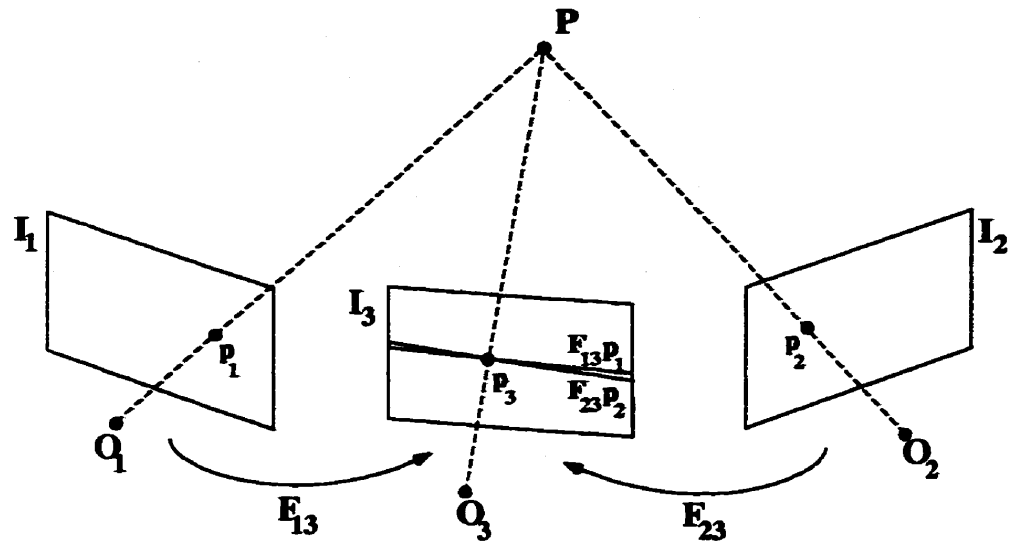


Figure 4.2 : Intersection des lignes épipolaires près du plan trifocal

4.1.4.3 Utilisation de la trilinearité

Les relations trilinéaires (Torr et Zisserman, 1997; Shashua, 1994b; Avidan et Shashua, 1997; Hartley, 1997c; Shashua et Werman, 1995) sont les relations qui lient les coordonnées (x_1, y_1) , (x_2, y_2) et (x_3, y_3) d'un triplet de points (p_1, p_2, p_3) en correspondance dans trois images (voir figure 4.3). Ces relations sont données par les équations suivantes :

$$\left\{ \begin{array}{l}
 a_1 + a_2x_1 + a_3x_3 + a_4y_1 + a_5y_2 + a_6x_1x_3 + a_7y_1y_2 + \\
 \quad a_8x_1y_2 + a_9x_3y_1 + a_{10}x_3y_2 + a_{11}x_3y_1y_2 + a_{12}x_1x_3y_2 = 0, \\
 a_{13} + a_{14}x_1 + a_{15}y_1 + a_{16}y_2 + a_3y_3 + a_{17}y_1y_2 + a_9y_1y_3 + \\
 \quad a_{10}y_2y_3 + a_{18}x_1y_2 + a_6x_1y_3 + a_{12}x_1y_2y_3 + a_{11}y_1y_2y_3 = 0, \\
 a_{19} + a_{20}x_1 + a_5x_2 + a_{21}x_3 + a_{22}y_1 + a_8x_1x_2 + a_{23}x_1x_3 + \\
 \quad a_{10}x_2x_3 + a_7x_2y_1 + a_{24}x_3y_1 + a_{11}x_2x_3y_1 + a_{12}x_1x_2x_3 = 0, \\
 a_{25} + a_{26}x_1 + a_{16}x_2 + a_{27}y_1 + a_{21}y_3 + a_{18}x_1x_2 + a_{24}y_1y_3 + \\
 \quad a_{17}x_2y_1 + a_{23}x_1y_3 + a_{10}x_2y_3 + a_{12}x_1x_2y_3 + a_{11}x_2y_1y_3 = 0.
 \end{array} \right.$$

La géométrie relative des trois caméras est représentée par les vingt-sept coefficients dénotés a_i . Ils forment un tenseur $3 \times 3 \times 3$ appelé *tenseur trilinéaire* qui est l'équivalent pour une configuration de 3 caméras ce qui est la matrice fondamentale pour une configuration de deux caméras. Comme les coefficients de la matrice fondamentale, ils sont liés par une condition de rang. Le positionnement relatif de 3 caméras perspectives peut être décrit par dix-huit paramètres. Donc, les coefficients du tenseur ne sont pas indépendants. Le tenseur trilinéaire peut être calculé par les mêmes méthodes que la matrice fondamentale.

Nous pouvons utiliser seulement les deux premières équations pour calculer la position du point (x_3, y_3) , connaissant le tenseur et les positions de (x_1, y_1) et (x_2, y_2) dans les deux premières images. Toutefois la position relative des images peut pro-

duire de mauvais résultats. Il faut donc considérer les différentes équations possibles (Shashua, 1997).

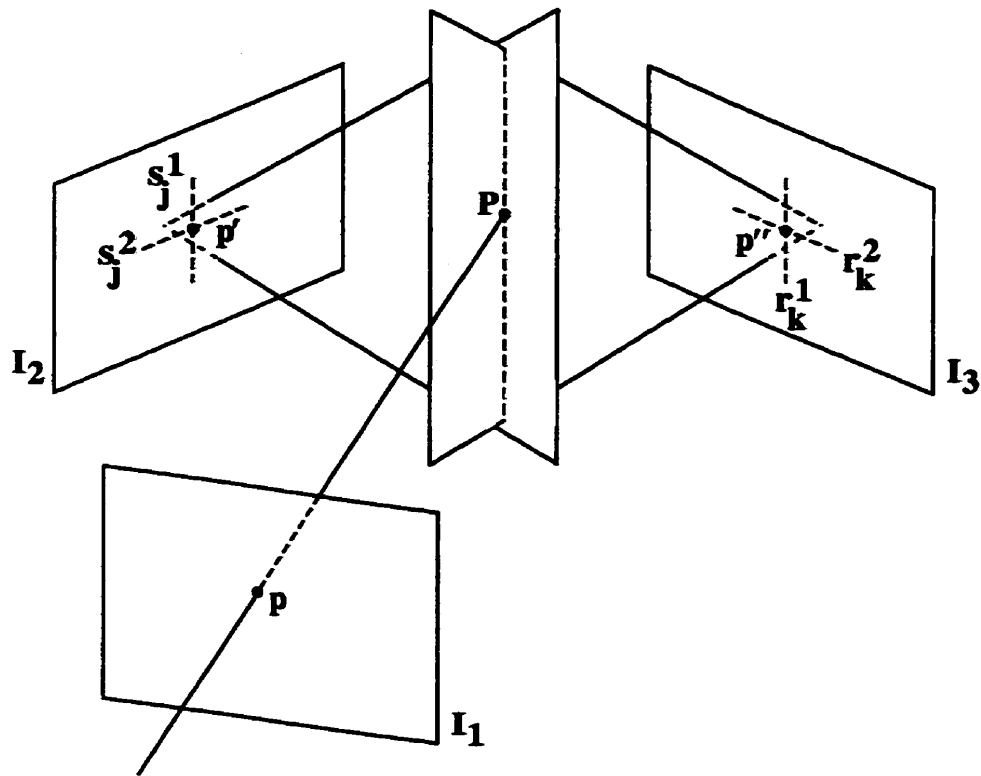


Figure 4.3 : Contraintes de trilinearité

4.1.5 Mosaïquage bidimensionnel

Ces dernières années un grand effort a été investi dans le domaine de la construction de mosaïques bidimensionnelles à partir d'images de référence. Plusieurs chercheurs se sont penchés sur cet aspect de la représentation de la scène qui permet de produire un large champ de vue à partir du collage des images de référence (Irani, Anandan et Hsu, 1995; Mase, 1996; Szeliski, 1996; Peleg et Herman, 1997; Rousso, Peleg, Finci

et Rav-Acha, 1998; Shum et Szeliski, 1998; Zoghلامي, Faugeras et Deriche, 1997). La représentation la plus utilisée est celle de la projection dans un repère cylindrique. Cette représentation est très adéquate lorsque les caméras décrivent un mouvement panoramique.

Chen et al. (Chen et Williams, 1993) proposent une représentation en mosaïque cylindrique dans le cas particulier où la caméra subit une rotation selon un axe vertical passant par son centre optique. Dans ce cas aucune information sur la structure 3D ne peut être récupérée. Les images sont liées par une homographie planaire qui peut être calculée à partir d'un appariement entre quatre points sur deux images, puis les images sont reprojctées sur un même repère cylindrique. Il est important d'avoir une zone de recouvrement entre les images qui se suivent pour permettre leur collage et l'établissement d'appariements entre au moins quatre points pour calculer l'homographie.

La compagnie Apple Computer commercialise un produit qui permet de produire une telle représentation et de visualiser interactivement le panorama généré de différents points de vues : *QuickTime VR*¹. Le principe de cette technique est illustré à la figure 4.4.

¹<http://www.apple.com/quicktime/qtvr/>

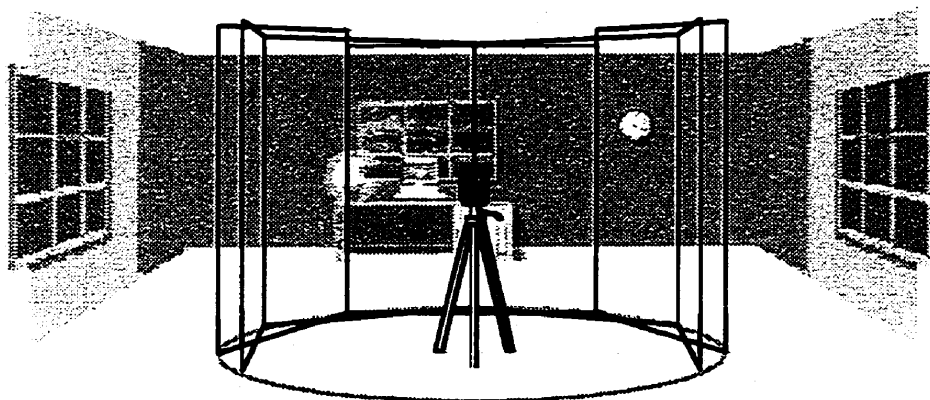


Figure 4.4 : Mosaïquage bidimensionnel cylindrique

4.1.6 Approche utilisant une base de données d'images

L'approche la plus connue dans ce contexte est celle de Levoy et al. (Levoy et Hanrahan, 1996) qui considèrent la scène comme étant décrite par un *champ lumineux*² : l'ensemble des rayons lumineux traversant la scène. Il est représenté par une fonction f à quatre dimensions, donnant la radiance en fonction de la position et de la direction d'observation. Levoy et al. utilisent, par exemple, un système de prises de vues constitué par une table tournante sur laquelle on pose l'objet. La table est synchronisée avec un système d'éclairage tournant et une caméra dont on contrôle le mouvement. Avec ce système on capture un grand nombre d'images (jusqu'à 8000 images). Tous les rayons sont décrits par la fonction f . Les images sont des échantillons bidimensionnels de cette fonction. Un grand nombre d'images permet de bien décrire f . Synthétiser une nouvelle vue revient alors à découper un échantillon bidimensionnel

²light field

dans la fonction f .

Etant donnée la quantité d'images disponibles les résultats obtenus sont de bonne qualité, toutefois la technique ne peut être utilisée que pour des objets de taille raisonnable.

4.2 Approche proposée

Nous proposons maintenant une nouvelle méthode de synthèse de vues intermédiaires à partir d'une paire d'images stéréoscopiques (Akhloufi, Cohen et Polotski, 1999; Akhloufi, Polotski et Cohen, 1999). Elle consiste tout d'abord à transférer les primitives d'intérêt dans la nouvelle vue. Pour cela, nous utiliserons la géométrie épipolaire ou les invariants projectifs pour réussir un transfert *géométriquement valide*. Une fois les primitives d'intérêt transférées dans la nouvelle vue, nous utiliserons une technique de *texturage bidimensionnel* (Heckbert, 1986; Wolberg, 1990) pour transférer les autres points dans cette même vue. Nous utiliserons une technique de texturage bidimensionnel perspective pour produire l'effet d'une projection centrale. Dans les sections suivantes nous allons présenter plus en détail les différents outils que nous avons utilisés pour la synthèse de vues. Nous finirons par une description de l'algorithme proposé.

4.3 Techniques de transfert de primitives d'intérêt

Dans la littérature on trouve principalement trois techniques de transfert de primitives d'intérêt qui permettent une reprojection *géométriquement valide* des points images dans la nouvelle vue : la géométrie épipolaire (Laveau et Faugeras, 1994), les trilinearités (Avidan, Evgeniou, Shashua et Poggio, 1997) et les invariants projectifs (Shashua, 1995). Ici nous allons nous intéresser plus particulièrement à la géométrie épipolaire et aux invariants projectifs. Nous allons souligner en particulier les possibilités offertes par la combinaison des deux méthodes pour transférer des points images dans une nouvelle vue.

4.3.1 Transfert à l'aide de la géométrie épipolaire

L'idée ici est de prédire la position d'une primitive d'intérêt dans une troisième vue à partir de ses correspondants dans deux images initiales (Faugeras et Robert, 1996). Pour cela nous allons utiliser les propriétés inhérentes à la géométrie entre paires de caméras ou géométrie épipolaire (voir chapitre 2). La géométrie de trois caméras est illustrée à la figure 4.5.

4.3.1.1 Transfert de points

Soient trois caméras (figure 4.5) dénotées par 1, 2 et 3, nous avons donc trois matrices fondamentales, F_{ij} , avec les indices $i, j = 1, \dots, 3, i \neq j$ représentant les caméras. Si m_i est un point image du plan image i , alors sa ligne épipolaire dans l'image j est

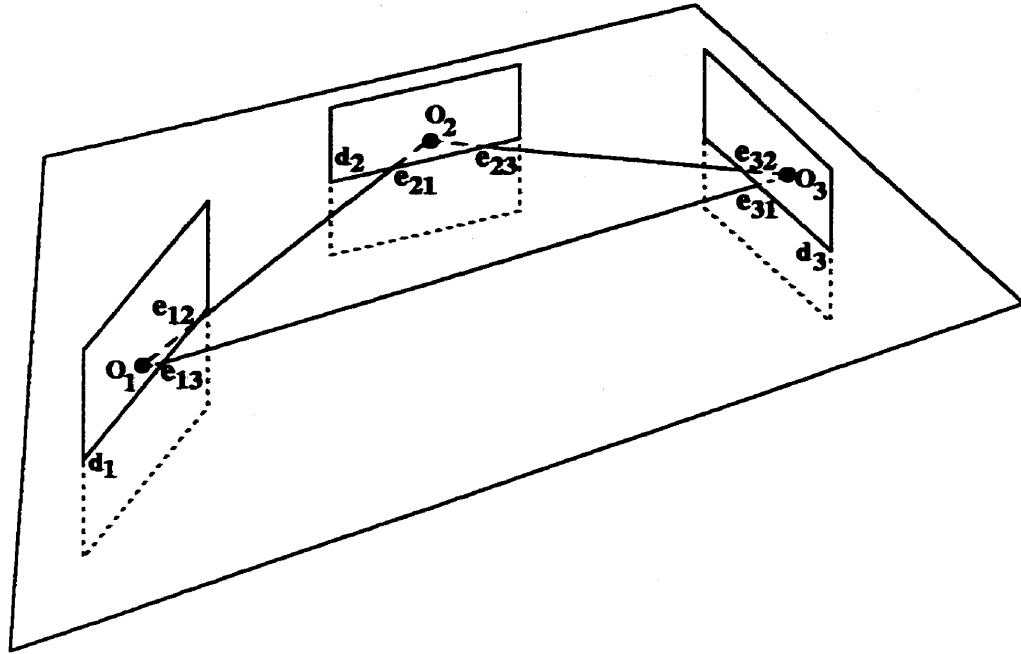


Figure 4.5 : Géométrie de trois caméras dans l'espace

représentée par ${}^i l = F_{ij} m_i$. On notera qu'on a la relation $F_{ij} = F_{ji}^t$. Le plan contenant les 3 centres optiques est le plan trifocal. Ce plan intersecte chaque plan image en une ligne notée d_i qui contient les épipoles $e_{i,i+1}$ et $e_{i,i+2}$ de la caméra i par rapport aux caméras $i+1$ et $i+2$. La géométrie épipolaire implique la relation suivante pour chaque image i :

$$F_{ii+1} e_{i,i+2} = d_{i+1} = e_{i+1,i} \otimes e_{i+1,i+2}.$$

Nous allons considérer le cas de transfert de points et de lignes dans une nouvelle vue.

Soient deux points appariés, m_1 et m_2 , dans les images 1 et 2. Donc, m_3 qui

leur correspond dans la troisième image doit se trouver nécessairement sur la ligne épipolaire de m_1 dans l'image 3, donnée par $F_{13}m_1$, et aussi sur la ligne épipolaire de m_2 dans cette troisième image, donnée par $F_{23}m_2$. Par conséquent m_3 se trouve à l'intersection des deux lignes épipolaires engendrées par les deux points appariés dans les images sources (voir figure 4.6) :

$$m_3 = F_{13}m_1 \otimes F_{23}m_2. \quad (4.3)$$

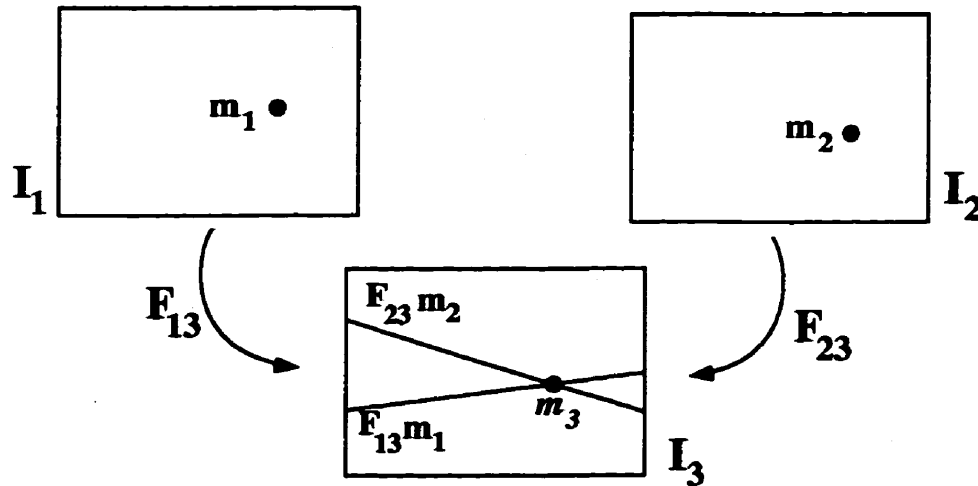


Figure 4.6 : Transfert de points à l'aide de la géométrie épipolaire

4.3.1.2 Transfert de lignes

La prédiction des lignes est obtenue d'une manière équivalente. Soient deux lignes appariées l_1 et l_2 dans les images 1 et 2 respectivement. Le but est de déterminer la position de la ligne l_3 , correspondant à l_1 et l_2 , dans l'image 3. Soient m_1 et m'_1 , deux points de l_1 . Ils définissent deux points m_2 , m'_2 de l_2 qui sont les intersections

des lignes épipolaires de m_1 représentées par $F_{12}m_1$ et de m'_1 représentée par $F_{12}m'_1$ avec la ligne l_2 . Nous avons

$$m_2 = F_{12}m_1 \otimes l_2,$$

et

$$m'_2 = F_{12}m'_1 \otimes l_2.$$

Alors la ligne l_3 est définie par deux points, m_3 et m'_3 , intersections des lignes épipolaires de m_1 et m_2 et de m'_1 et m'_2 dans la troisième image (voir figure 4.7) :

$$l_3 = m_3 \otimes m'_3$$

ou bien

$$l_3 = (F_{13}m_1 \otimes F_{23}m_2) \otimes (F_{13}m'_1 \otimes F_{23}m'_2). \quad (4.4)$$

4.3.1.3 Transfert proche du plan trifocal

On peut facilement constater que les méthodes décrites plus haut ne sont plus applicables lorsque les droites épipolaires engendrées dans la troisième image sont confondues. En effet, sur le plan trifocal la droite épipolaire engendrée par m_1 est confondue avec celle engendrée par m_2 . Nous allons décrire dans la suite une méthode permettant de résoudre ce problème (Laveau et Faugeras, 1994).

Soient deux paires de lignes épipolaires choisies loin de l'intersection du plan tri-

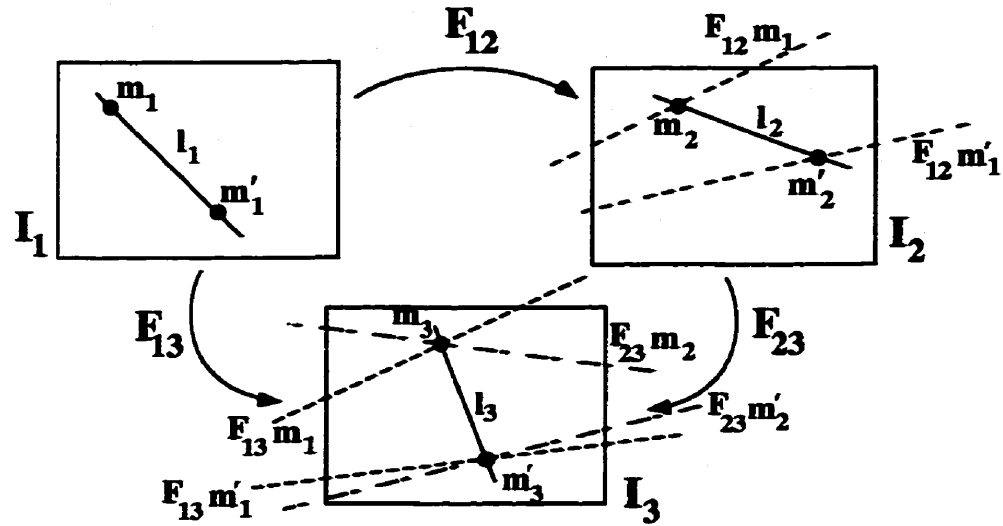


Figure 4.7 : Transfert de lignes à l'aide de la géométrie épipolaire

focal avec les plans images. Nous les notons (l_1^p, l_2^p) et (l_1^q, l_2^q) . Soient p_1 , un point choisi sur l_1^p , et p_2 , un point choisi sur l_2^p , représentant un point P de la scène. Si M est un point de l'espace appartenant au plan trifocal correspondant à m_1 et m_2 respectivement sur l'image 1 et 2, la ligne (PM) rencontre le plan défini par (l_1^q, l_2^q) en un point noté Q . Soient q_1 et q_2 les images de Q dans les images 1 et 2. Nous construisons les points p_3 et q_3 comme l'intersection des lignes épipolaires engendrées par p_1 et p_2 respectivement q_1 et q_2 . Comme P et Q sont loin du plan trifocal alors p_3 et q_3 sont bien définis (voir figure 4.8).

Nous pouvons déduire m_3 comme l'intersection de la ligne passant par les épiholes e_{31} et e_{32} , définie par $e_{31} \otimes e_{32}$, et de la ligne définie par $p_3 \otimes q_3$ (voir figure 4.9).

Pour éviter les problèmes liés à la reprojection des points images appartenant au plan trifocal dans la nouvelle vue, nous présentons une autre méthode dans la section

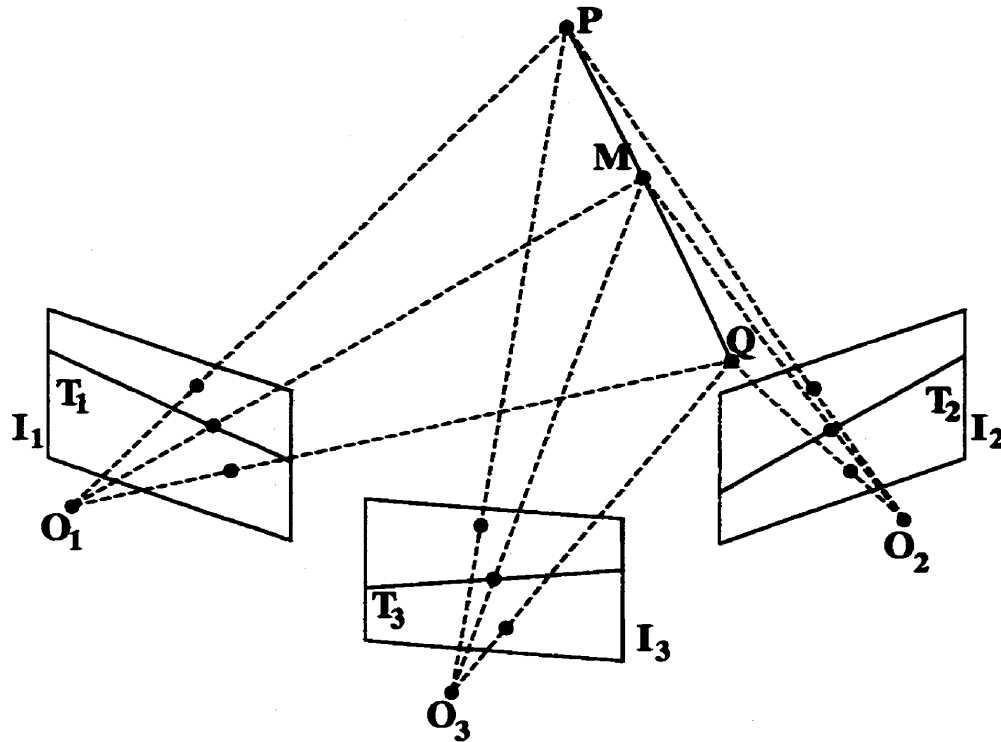


Figure 4.8 : Transfert proche du plan trifocal: choix des points P et Q loin du plan trifocal

suivante. Cette méthode est basée sur l'invariance projective chapitre 3.

4.3.2 Transfert à l'aide d'un invariant projectif

Nous avons vu au chapitre précédent (chapitre 3) que la théorie d'invariance nous offre des outils permettant de définir une propriété invariante d'une configuration géométrique présente dans la scène. Nous allons montrer comment nous pouvons utiliser les invariants pour transférer les primitives d'intérêt dans la nouvelle vue.

Dans la section 3.2 nous avons défini en termes de birapports un invariant projectif à partir de deux homographies caractérisant deux plans de la scène et un point de

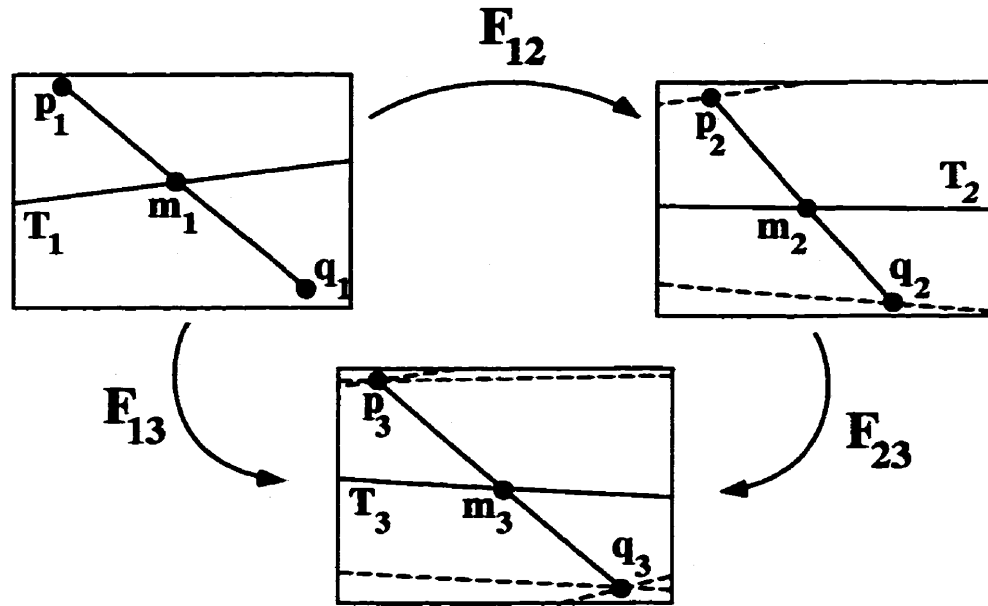


Figure 4.9 : Transfert proche du plan trifocal à l'aide de la géométrie épipolaire

référence :

$$\bar{k} = [P_1, P_o, P_x, P_2] = [p_1^1, p_o^1, p_x^1, p_2^1] = [p_1^2, p_o^2, p_x^2, p_2^2]. \quad (4.5)$$

L'invariance pour toute transformation projective implique que \bar{k} reste le même pour les points correspondants sur toutes les images. Ainsi pour trois points correspondants, p_1 , p_2 et p_3 , se trouvant respectivement sur les plans images 1, 2 et 3, nous avons le même invariant projectif défini par \bar{k} . Nous utiliserons cette propriété pour le transfert des points images dans une troisième vue. Pour cela, nous avons besoin de reprojeter tout d'abord les deux plans de référence (Π_1 et Π_2) et le point de référence (P_o) dans la troisième vue. A ce stade nous utilisons la *géométrie épipolaire* pour y arriver. Quatre points (au moins) sont identifiés dans chacun des plans de référence, puis transférés sur l'image 3 en utilisant l'intersection des lignes épipolaires définies

par les matrices fondamentales F_{13} et F_{23} entre les images 1 et 3 respectivement 2 et 3, comme décrit dans la section plus haut sur le transfert à l'aide de la géométrie épipolaire. Ceci nous permet de définir l'homographie caractéristique de chaque plan. De la même façon, le point de référence sur la troisième vue est obtenu par l'intersection des lignes épipolaires engendrées par ses correspondants dans les vues 1 et 2 (équation 4.3) (voir figure 4.10).

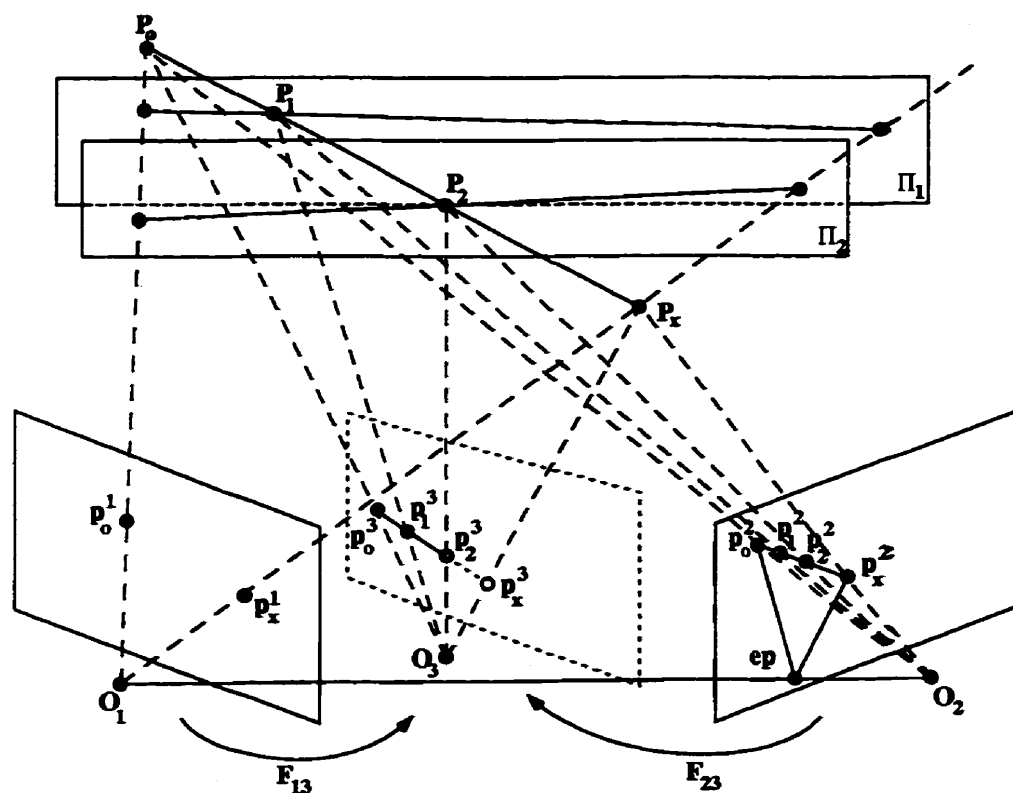


Figure 4.10 : Transfert à l'aide d'un invariant projectif

Une fois que les positions des deux plans de référence et du point de référence sur la troisième image sont calculées, on peut déduire tous les points dans cette nouvelle vue à partir de leurs correspondants et des invariants projectifs qui les caractérisent.

Soit un point quelconque P_x , caractérisé par un invariant noté \bar{k} , on a :

$$\bar{k} = [p_1^1, p_o^1, p_x^1, p_2^1] = [p_1^2, p_o^2, p_x^2, p_2^2] = [p_1^3, p_o^3, p_x^3, p_2^3].$$

Donc pour calculer p_x^3 (le point image de P_x sur la nouvelle vue dénotée par 3), il suffit d'avoir p_1^3 , p_o^3 et p_2^3 (l'invariant \bar{k} étant calculé à partir des deux images sources). p_o^3 est la reprojection du point de référence que nous obtenons à l'aide de la géométrie épipolaire. p_1^3 et p_2^3 peuvent être obtenus à partir des homographies entre les plans de référence entre les images 1 et 3 ou bien les images 2 et 3 (section 3.2).

4.4 Texturage bidimensionnel

En infographie les techniques classiques de calquage de textures permettent l'adjonction de motifs (images bidimensionnelles) à des objets tridimensionnels, pour arriver à effectuer une synthèse d'images réalistes (Foley et al., 1990; Watt, 1993). Le rendu³ réaliste a pour but de produire des images non discernables de photographies d'objets réels. Ces techniques sont très utilisées en infographie où on dispose d'un modèle tridimensionnel de la scène. Lorsque l'information sur la structure de la scène est absente nous pouvons utiliser des techniques de *texturage bidimensionnel* (Wolberg, 1990) qui permettent la reprojection de textures d'un espace bidimensionnel dans un autre espace bidimensionnel. Ainsi le texturage bidimensionnel est un ensemble de

³Rendering

transformations spatiales, auxquelles on ajoute des techniques de filtrage de l'image pour éliminer les bruits induits par ces transformations. Pour produire un effet visuel acceptable et éliminer les bruits inhérents aux transformations géométriques nous utilisons les mêmes techniques qu'en rendu classique: rééchantillonnage, anti-crênelage⁴, etc.

Toute transformation géométrique est représentée par la projection d'un système de coordonnées vers un autre système de coordonnées. Elle est définie par des transformations spatiales: une application qui établit la correspondance spatiale entre tous les points de l'image source et de l'image destination. La plupart des transformations peuvent être décrites par des expressions analytiques simples, comme dans le cas des transformations affines, projectives, bilinéaires et polynomiales. D'autres transformations plus complexes peuvent être déterminées par un ensemble de points de contrôle éparses dont la correspondance sur les deux images est connue. Ceci permet de définir une représentation spatiale dans laquelle on détermine les points inconnus par interpolation entre les points de contrôle. Dans la suite, nous nous intéresserons plus particulièrement aux transformations perspectives. Nous allons introduire aussi les transformations affines et bilinéaires qui présentent un intérêt indéniable en texturage bidimensionnel.

⁴L'anti-crênelage désigne la technique qui porte le nom anglais de "*antialiasing*"

4.4.1 Transformation affine

Une transformation projective générale entre deux plans images est décrite par une matrice de transformation 3×3 notée T :

$$T = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

La transformation affine est caractérisée par une dernière colonne de la matrice T de la forme $[0, 0, 1]^t$, qui correspond à une *projection planaire parallèle* ou *orthographique*. La représentation générale d'une transformation affine est

$$[x, y, 1] = [u, v, 1] \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix}.$$

Ici, $[x, y, 1]$ sont les coordonnées homogènes d'un point de l'image destination et $[u, v, 1]$ les coordonnées homogènes d'un point de l'image source. Nous utiliserons cette notation dans la suite.

Les transformations affines préservent les lignes parallèles lors de la reprojection. Elles ont 6 degrés de liberté. Pour inférer une transformation affine, il suffit d'établir la correspondance entre (au moins) trois points dans l'image de destination et trois point dans l'image source. Les transformations affines permettent un ensemble de

transformations planaires limitées, par exemple la transformation d'un triangle vers un triangle quelconque ou d'un rectangle vers un parallélogramme. Des distorsions plus complexes nécessitent des transformations plus élaborées, comme les transformations perspectives ou bilinéaires.

4.4.2 Transformation perspective

La transformation perspective est caractérisée par $[a_{13}, a_{23}]^t$ non nul. La représentation générale de cette transformation est,

$$[x, y, 1] = [u, v, 1] \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

La transformation perspective est une transformation projective liée à une projection centrale. Elle préserve les lignes parallèles seulement lorsqu'elles sont parallèles au plan de projection, autrement elles convergent en un point de fuite. Cette technique est très utile pour produire des rendus réalistes d'images. La transformation perspective est équivalente à l'homographie en géométrie projective, que nous avons défini dans le chapitre 2. Cette transformation a huit degrés de liberté puisque un des paramètres de la matrice peut être fixé égal à 1. Pour la calculer nous avons besoin d'établir la correspondance entre (au moins) quatre points dans l'image de destination et l'image source. Lorsqu'il s'agit de transformer un quadrilatère dans un autre

quadrilatère communément appelée *projection des quatre coins*⁵, la transformation perspective nous offre une solution planaire à ce problème. Lorsque les quadrilatères ne sont pas planaires, il faut des solutions plus générales. La transformation bilinéaire est la transformation la plus simple qui peut produire une projection des quatre coins pour des quadrilatères non planaires.

4.4.3 Transformation bilinéaire

La transformation bilinéaire est donnée par

$$[x, y] = [uv, u, v, 1] \begin{bmatrix} a_0 & b_0 \\ a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{bmatrix}.$$

Cette transformation permet de produire une projection entre quadrilatères non plans. Elle préserve les lignes se trouvant sur la direction de balayage dans l'image source. Cette propriété vient de la nature des interpolations bilinéaires utilisées pour faire cette transformation. Ainsi par exemple, les points le long des lignes horizontales et verticales dans l'image source restent équidistants après transformation, comme en projection affine. Toutefois les lignes qui ne sont pas alignées selon ces deux directions ne sont pas préservées. Par exemple, les lignes diagonales sont transformées en des

⁵Nous avons utilisé la *projection des quatre coins* pour désigner la technique connue sous le nom de *four-corner mapping* en infographie.

courbes quadratiques. Cette technique est largement utilisée en calquage de textures en infographie. Ceci vient du fait que cette transformation peut être décomposée en deux transformations unidimensionnelles permettant un gain appréciable en temps de calcul.

4.4.4 Décomposition de l'image

Bien que les transformations bilinéaires permettent des transformations plus générales, entre quadrilatères par exemple, elles introduisent des distorsions lorsqu'il s'agit de surfaces planes. Pour produire des effets plus réalistes, il est préférable de décomposer l'image source en des surfaces les plus proches possible de plans puis utiliser une transformation perspective pour projeter ces surfaces sur l'image de destination. Nous avons besoin donc d'établir la correspondance entre quatre points dans les deux images. Ceci signifie que nous avons besoin de décomposer l'image source en une grille formée de quadrilatères, puis déterminer la transformation perspective permettant la reprojection entre chaque paire de quadrilatères appariés. Toutefois la méthode peut être étendue aux cas de surfaces triangulaires en utilisant la géométrie épipolaire. Soit l'image source décomposée en des surfaces triangulaires (en utilisant une triangulation de Delaunay par exemple (Le Bras-Mehlman, Schmitt, Boissonnat et Faugeras, 1988)). En établissant l'appariement entre surfaces triangulaires dans les images source et destination, nous obtenons une paire de trois points appariés pour chaque surface. Ceci caractérise une transformation affine. Pour définir une

transformation perspective il suffit d'identifier une quatrième paire de points qui satisfait la même transformation perspective que les trois autres, et ceci pour chaque surface triangulaire de l'image. La géométrie épipolaire nous permet de définir toute transformation perspective ou homographique en terme des épipoles. Ainsi, quelque soit la transformation perspective permettant la reprojection d'une surface plane de l'image source en une surface plane de l'image destination, elle transforme l'épipole dans l'image source vers l'épipole dans l'image destination (Shashua, 1994a). Nous avons

$$[x_i, y_i, 1] = [u_i, v_i, 1] \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad (4.6)$$

où $i = 1, 2, 3$ représente les trois points de la surface triangulaire, et

$$[ep_{1d}, ep_{2d}, 1] = [ep_{1s}, ep_{2s}, 1] \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad (4.7)$$

où $[ep_{1d}, ep_{2d}, 1]$ et $[ep_{1s}, ep_{2s}, 1]$ définissent les coordonnées homogènes des épipoles dans les images de destination respectivement source.

Donc, en résolvant le système défini par les équation 4.6 et 4.7, nous obtenons la matrice de transformation perspective. Ainsi, à partir de la triangulation de l'image source et l'appariement avec des triangles dans l'image de destination nous pou-

vons produire une transformation perspective entre chacune des surfaces triangulaires. Dans la littérature on trouve plusieurs techniques de triangulation, dont la plus utilisée est la triangulation de Delaunay.

4.4.4.1 Anti-crénelage

L'anti-crénelage (Wolberg, 1990; Foley et al., 1990) est un outil très important en infographie. Il permet de réduire l'effet de certains types de bruit introduits lors des transformations d'image. Des artefacts apparaissent dans l'image lorsque les transformations induisent des changements à grande échelle. Lorsqu'on augmente l'échelle d'une zone d'image chaque pixel source contribue à plusieurs pixels destinations. Ce type de transformation nécessite un échantillonnage dense du signal reconstruit. La qualité de l'image résultante est liée à la précision des fonctions d'interpolation utilisées lors de la reconstruction. Ainsi, des fonctions d'interpolation de degré supérieur peuvent produire de meilleures reconstructions d'un grand nombre de signaux, contrairement à des fonctions d'interpolation de degré inférieur.

Il y a deux méthodes principales d'anti-crénelage. La première consiste à accroître le taux d'échantillonnage. Cela s'obtient en augmentant la résolution de la trame. Des détails plus fins sont ainsi incorporés. La seconde méthode consiste à traiter un pixel comme une zone finie et non pas comme un point. Ceci est équivalent à un préfiltrage de l'image.

4.5 Algorithme de synthèse de vues

En résumé l'algorithme proposé pour la synthèse de vues se divise principalement en quatre étapes (voir figure 4.11) :

1. Sélection des primitives d'intérêt appariées (points, lignes) sur les deux images sources.
2. Décomposition de l'image (section 4.4.4) en utilisant les n plus proches voisins : $n = 4$ pour un maillage en quadrilatères et $n = 3$ pour une triangulation (comme nous avons vu dans la section 4.4.4, plus les surfaces résultants de la décomposition sont proches de l'image de structures physiques planes, plus le texturage bidimensionnel perspective sera réaliste).
3. Transfert des primitives d'intérêt sur la nouvelle vue (par les techniques décrites dans la section 4.3).
4. Texturage bidimensionnel (section 4.4) entre les surfaces correspondantes dans l'image source (la plus proche de la vue à générer) et la nouvelle image.

4.6 Conclusion

La synthèse de vues est un domaine récent et est en évolution constante. Ces dernières années les techniques sont devenues plus matures, et de nombreuses publications apparaissent chaque année. Dans ce chapitre nous avons commencé par présenter l'état

de l'art en synthèse de vues où nous avons passé en revue les principales techniques utilisées que nous avons divisées en trois grandes classes: reconstruction tridimensionnelle explicite, rendu basé sur image, et mosaïquage ou représentation étendue.

Dans la deuxième partie, nous avons proposé une nouvelle approche qui permet la génération de nouvelles vues à partir seulement d'un petit nombre d'appariements dans l'image. L'appariement épars est utilisé pour générer un ensemble de points qui vont jouer le rôle de points de contrôle. Nous utilisons ces points pour définir un maillage sur l'image source, qui est reprojété sur la nouvelle vue à l'aide de la géométrie épipolaire et de l'invariance projective. Ainsi, la vue générée est géométriquement valide. Nous avons utilisé des techniques de texturage bidimensionnel pour reprojeter le reste des points de l'image. Nous avons choisi une technique de texturage perspective qui nécessite d'utiliser au moins un maillage en quadrilatères. Lorsque l'image est décomposée en surfaces triangulaires, nous avons introduit une technique permettant l'utilisation des contraintes définies par la géométrie épipolaire pour produire un texturage perspective valide. Nous avons fini ce chapitre avec une description globale de l'algorithme proposé.

Contrairement aux approches existantes, nous n'avons besoin que d'un appariement épars pour générer la nouvelle vue, ce qui réduit la complexité associée aux techniques d'appariements. Ainsi le temps de calcul nécessaire à l'appariement est largement réduit. De même la synthèse de vues à partir d'images connues permet d'éviter les étapes de modélisation et de rendu d'une scène tridimensionnelle qui impliquent sou-

vent des coûts élevés en temps machine. L'application des outils de la géométrie projective nous permet de produire des vues assez réalistes, puisqu'elles satisfont les contraintes géométriques inhérentes à la scène.

Dans le chapitre suivant nous montrerons les résultats obtenus par cette technique de synthèse de vues.

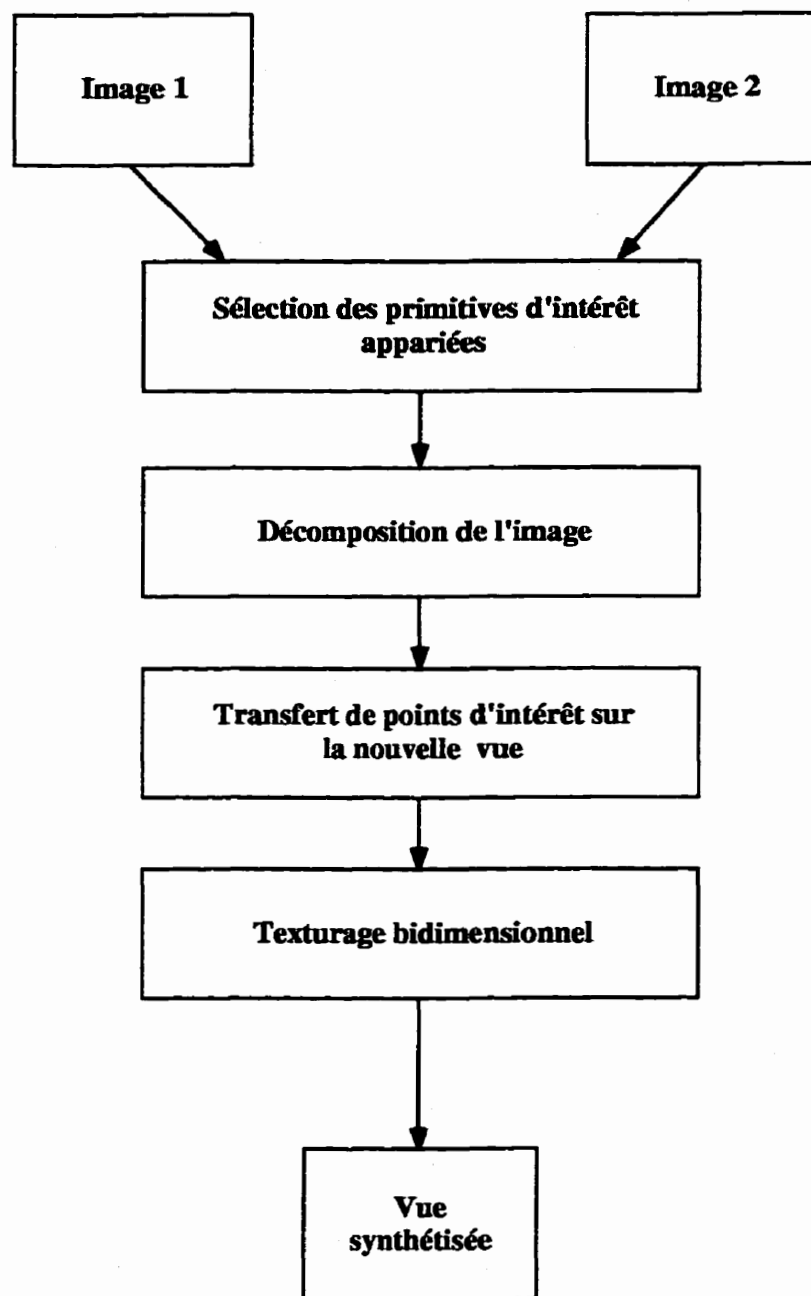


Figure 4.11 : Description de l'approche proposée

Chapitre 5

Résultats expérimentaux

Dans ce chapitre nous allons présenter les résultats obtenus par les différentes méthodes décrites dans les chapitres précédents. Nous allons tout d'abord présenter les résultats de la géométrie épipolaire sur des paires d'images stéréoscopiques¹. Ensuite, nous allons présenter les résultats pour les invariants projectifs. Enfin nous montrerons une application particulière des résultats précédents: la synthèse de nouvelles vues.

5.1 Géométrie épipolaire

Dans cette section nous nous intéressons à l'application de la méthode basée sur l'homographie que nous avons développée dans la section 2.5 pour calculer la matrice

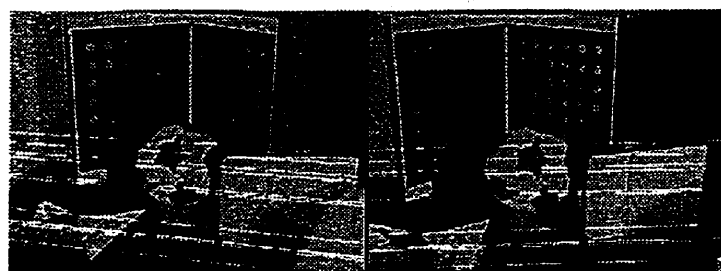
¹Les images stéréoscopiques utilisées sont disponibles publiquement à l'adresse du groupe SYNTIM à: <http://www-syntim.inria.fr/syntim/analyse/paires-eng.html>, les autres images ont été générées au laboratoire du Groupe de Recherche en Perception et Robotique, École Polytechnique de Montréal.

fondamentale et sa comparaison avec différentes méthodes existantes dans la littérature. Les méthodes que nous utilisons dans nos comparaisons ont été décrits dans la section 2.4. Toutes ces méthodes sont comparées à la technique de calibration classique (Tsai, 1987; Zhang, 1996; Torr et Murray, 1997) qui est considérée comme étant le résultat qui représente le mieux la géométrie inhérente à une configuration de caméras donnée (en calibration classique l'information de profondeur est utilisée pour obtenir des résultats plus fidèles à la réalité).

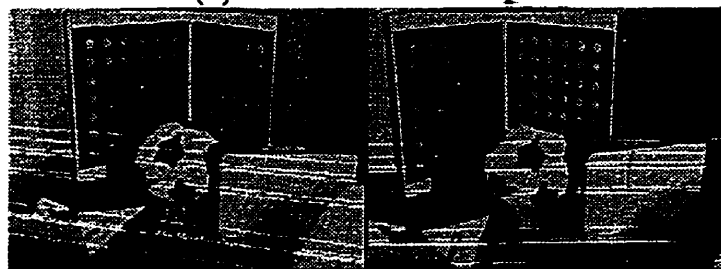
5.1.1 Matrice fondamentale

Nous utilisons les principales techniques décrites dans les sections 2.4 et 2.5 pour calculer la matrice fondamentale. Les résultats de la géométrie épipolaire estimée sont montrés sur les figures 5.1, 5.2, 5.3, 5.4 et 5.5.

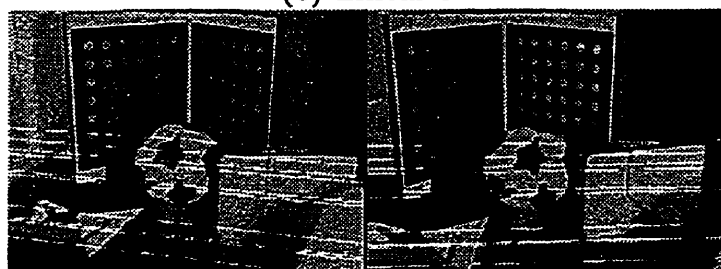
On remarque que les lignes épipolaires obtenues par la matrice fondamentale calculée par l'approche que nous avons proposée et qui est basée sur l'homographie (F_h) sont très proches de celles obtenues par la méthode de moindre médiane des carrés (F_{lms}) qui est connue pour être la plus efficace des méthodes de calcul de matrice fondamentale (Torr et Murray, 1997; Zhang, 1996). Cette dernière, avec notre méthode, donne les résultats les plus proches de ceux du résultat de calibration stéréoscopique classique (F_{cal}) (voir figures 5.1, 5.2, 5.3, 5.4 et 5.5). La méthode linéaire (F_{lin}) donne le plus mauvais résultat. Mais après normalisation des données (F_{linN}) nous obtenons une nette amélioration pour cette dernière, les résultats sont proches des



(a) Calibration classique



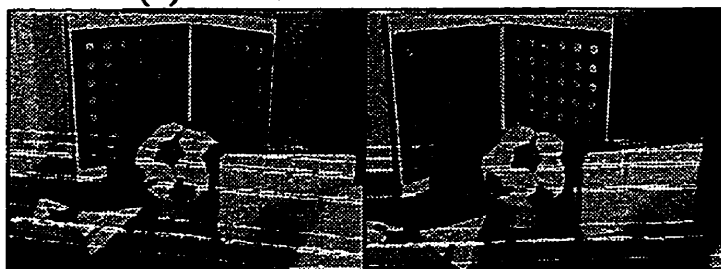
(b) Linéaire



(c) Non linéaire



(d) Moindre médiane des carrés

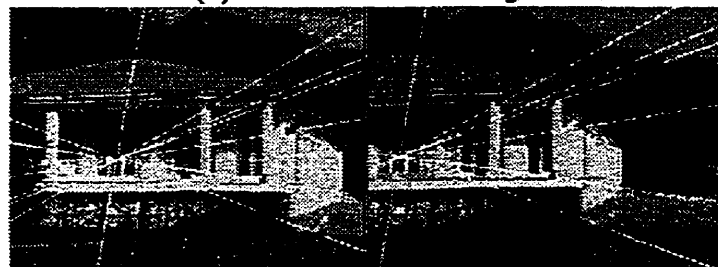


(e) Homographie

Figure 5.1 : Résultats des différentes techniques d'estimation de la géométrie épipolaire sur la paire stéréo "Mire"



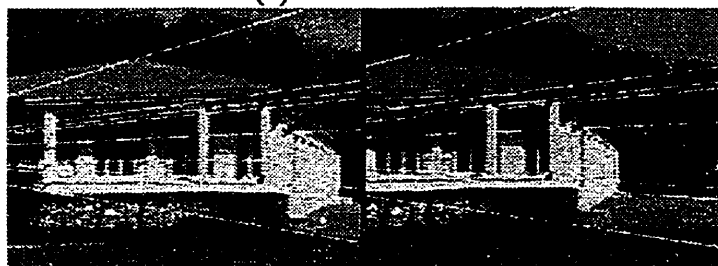
(a) Calibration classique



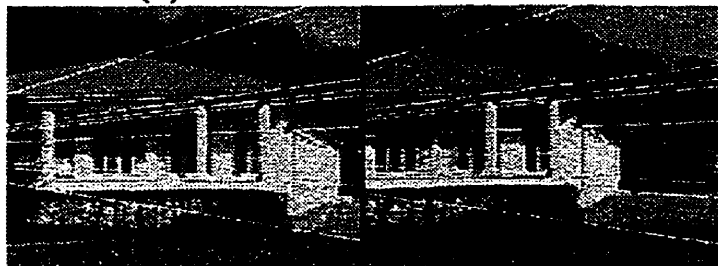
(b) Linéaire



(c) Non linéaire

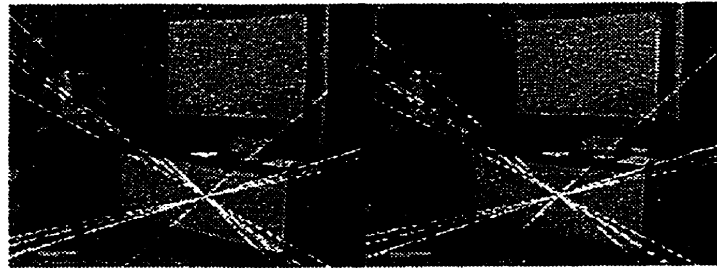


(d) Moindre médiane des carrés

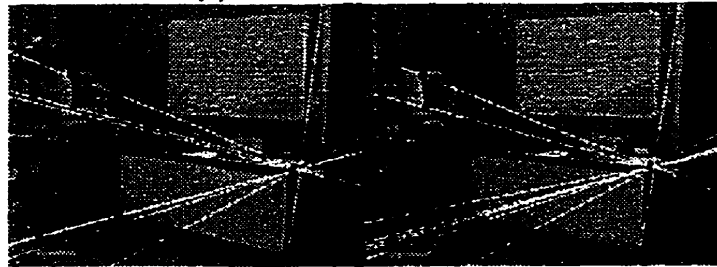


(e) Homographie

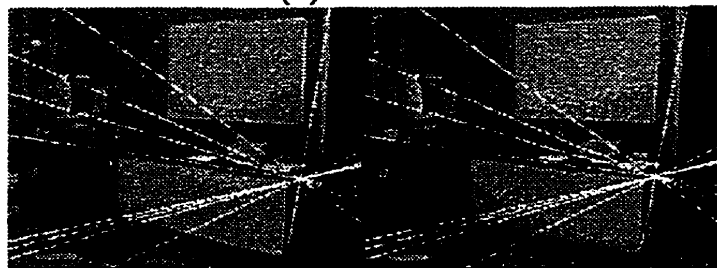
Figure 5.2 : Résultats des différentes techniques d'estimation de la géométrie épipolaire sur la paire stéréo "Bâtiment"



(a) Calibration classique



(b) Linéaire



(c) Non linéaire

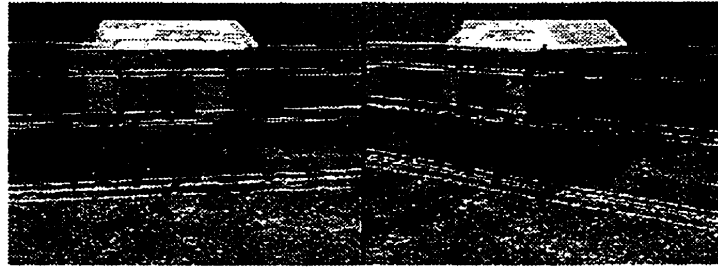


(d) Moindre médiane des carrés

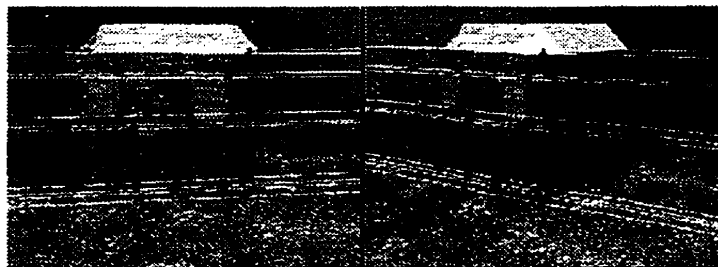


(e) Homographie

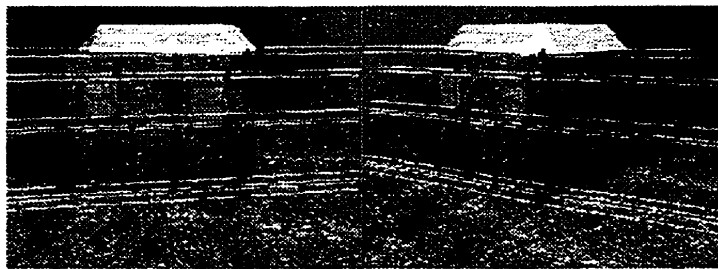
Figure 5.3 : Résultats des différentes techniques d'estimation de la géométrie épipolaire sur la paire stéréo "Bureau"



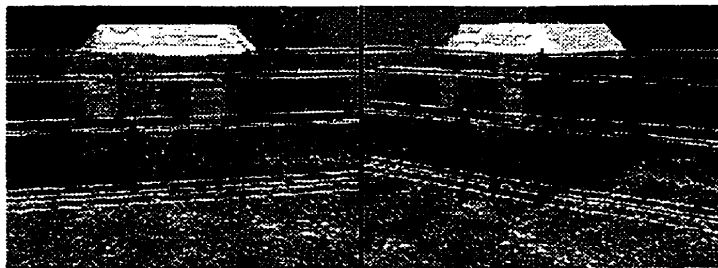
(a) Linéaire



(b) Non linéaire

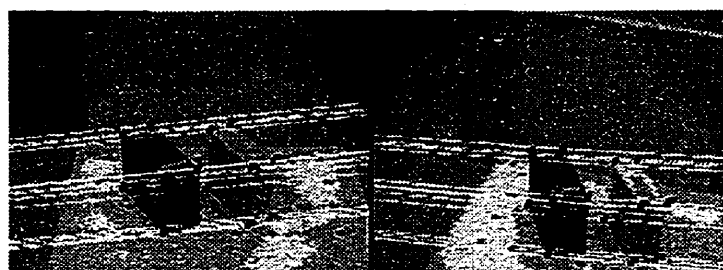


(c) Moindre médiane des carrés

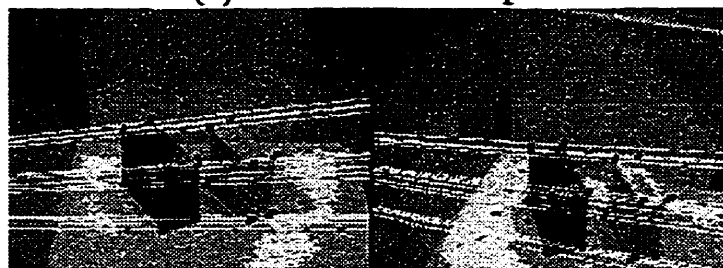


(d) Homographie

Figure 5.4 : Résultats des différentes techniques d'estimation de la géométrie épipolaire sur la paire stéréo "Maison"



(a) Calibration classique



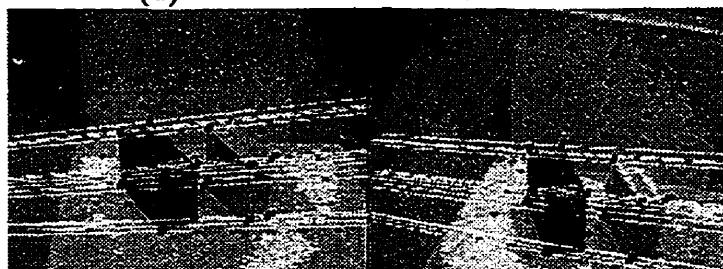
(b) Linéaire



(c) Non linéaire



(d) Moindre médiane des carrés



(e) Homographie

Figure 5.5 : Résultats des différentes techniques d'estimation de la géométrie épipolaire sur la paire stéréo "Lab"

résultats de la méthode non linéaire (F_{nlin}) qui reste toutefois loin de la méthode de moindre médiane des carrés et de la méthode basée sur l'homographie que nous avons développée.

Dans la suite nous allons présenter les résultats de comparaison entre différentes méthodes de calcul de la matrice fondamentale.

5.1.2 Comparaisons

Nous nous intéressons plus particulièrement à la comparaison de notre approche aux principales méthodes existantes.

Ces dernières années le problème de comparaison entre différentes méthodes de calcul de la matrice fondamentale a suscité beaucoup d'intérêt au sein de la communauté de vision par ordinateur. L'objectif est de trouver la meilleure mesure pour comparer entre deux matrices fondamentales. Par le passé la comparaison se faisait principalement en comparant visuellement les lignes épipolaires obtenues et superposées sur l'image correspondante, on vérifiait si les lignes passaient par les points correspondants et si elles étaient proches des lignes obtenues par les résultats de calibration stéréoscopique classique (Tsai, 1987). Une autre technique pour comparer entre les matrices consistait à calculer la norme de Fobenius de la différence entre deux matrices fondamentales normalisées, toutefois il a été démontré que cette dernière mesure ne permettait pas de vérifier l'efficacité d'un algorithme par rapport aux autres (Torr et Murray, 1997; Zhang, 1996). Ainsi de nouvelles techniques de

mesures furent introduites pour permettre de calculer des quantités mesurables directement sur l'image. Deux méthodes sont utilisées dans ce sens. La première est appelée mesure de différence entre matrices fondamentales notée **Fdiff** (Csurka et al., 1997). Elle permet de comparer entre les matrices fondamentales en terme de distance entre points et lignes épipolaires correspondantes en utilisant simultanément les deux matrices fondamentales. La deuxième appelée facteur de qualité permet quand à elle la mesure de la distance moyenne entre les lignes épipolaires calculées et les points utilisés pour calculer la matrice fondamentale, dénotée par Q_F (Boufama et Mohr, 1995).

5.1.2.1 Différence entre matrices fondamentales: Fdiff

Dans la suite nous décrivons comment obtenir la mesure de différence entre matrices fondamentales. Soient deux matrices fondamentales à comparer, F_1 et F_2 . La mesure de comparaison est donnée par la procédure suivante (figure 5.6) :

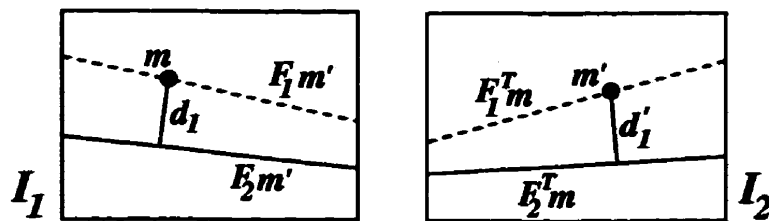


Figure 5.6 : Différence entre matrices fondamentales

1. Choisir de façon aléatoire un point m dans la première image.

2. Tracer la ligne épipolaire de m dans la deuxième image en utilisant F_1 . (ligne pointillée sur la figure 5.6 définie par $F_1^T m$).
3. Revenir à l'étape 1 si la ligne épipolaire ne croise pas la deuxième image.
4. Choisir de façon aléatoire un point m' sur la ligne épipolaire.
5. Tracer la ligne épipolaire de m dans la deuxième image en utilisant F_2 , soit $F_2^T m$, et calculer la distance d'_1 entre le point m' et la ligne $F_2^T m$.
6. Tracer la ligne épipolaire de m' dans la première image en utilisant F_2 , soit $F_2 m'$, et calculer la distance d_1 entre le point m et la ligne $F_2 m'$.
7. Refaire les étapes 2 à 6 en inversant les rôles de F_1 et F_2 et calculer d_2 et d'_2 .
8. Répéter N fois les étapes 1 à 7.
9. Calculer la distance moyenne en pixels des d_i et d'_i , qui représente la mesure de comparaison entre les deux matrices fondamentales.

Les résultats pour les images des figures 5.1, 5.2, 5.3, 5.4 et 5.5 sont donnés par les tableaux 5.1, les valeurs représentent les erreurs ou distances moyennes en pixels. Nous avons d'abord testé les différentes méthodes décrites dans les sections 2.4 et 2.5 sur l'image de la paire stéréoscopique "Mire", par la suite nous avons choisi les techniques les plus représentatives de chaque classe (classes des méthodes : linéaires, non linéaires, et robustes) pour poursuivre les tests de comparaison par rapport aux résultats de calibration classique sur les autres paires d'images.

Tableau 5.1 : Résultats de comparaison entre différentes techniques de calcul de la matrice fondamentale par la technique Fdiff

(a) Paire stéréo "Mire"

	Fcal	Flin	FlinN	Fh	Fnlín	FnlínN	Fgrad	Fng	Fmest	Fms
Fcal	0	124.16	3.87	2.11	3.41	4.46	3.42	4.44	1.56	0.77
Flin		0	124.06	124.72	123.70	124.67	123.46	124.96	123.63	124.03
FlinN			0	2.83	0.63	0.73	0.63	0.73	5.41	3.28
Fh				0	2.26	3.08	2.87	3.07	2.62	2.23
Fnlín					0	1.09	0.00	1.09	4.87	2.78
FnlínN						0	1.08	0.01	6.00	3.83
Fgrad							0	1.09	4.89	2.76
Fng								0	5.90	3.86
Fmest									0	2.10
Fms										0

(b) Paire stéréo "Bâtiment"

	Fcal	FlinN	Fh	Fnlín	Fms
Fcal	0	3.09	1.80	2.83	1.40

(c) Paire stéréo "Bureau"

	Fcal	FlinN	Fh	Fnlín	Fms
Fcal	0	4.76	2.71	4.13	1.50

(d) Paire stéréo "Lab"

	Fcal	FlinN	Fh	Fnlín	Fms
Fcal	0	4.27	1.62	3.77	1.17

Nous constatons que la méthode linéaire (F_{lin}) donne les plus mauvais résultats. La normalisation des données a priori (F_{linN}) permet d'améliorer les résultats de la méthode linéaire. La méthode non linéaire (F_{nlin}) donne de meilleurs résultats que la méthode linéaire. Ces mesures sont équivalentes à celles obtenues par la méthode de gradient (F_{grad}). La normalisation des données avant l'application des méthodes non linéaires et de gradient ne permet pas par contre l'amélioration des résultats (F_{nlinN} pour méthode non linéaire normalisée et F_{ng} pour méthode de gradient normalisée). Toutefois, les meilleurs résultats sont ceux obtenus par les méthodes non linéaires utilisant les techniques de statistiques robustes (M-estimateurs (F_{mest}) et moindre médiane des carrés (F_{lms})). La méthode linéaire basée sur l'homographie (F_h) que nous avons développée donne de meilleurs résultats que les autres méthodes linéaires existantes et s'approche des résultats obtenus par les méthodes non linéaires utilisant les statistiques robustes.

En somme, la méthode que nous avons proposée dans la section 2.5 est plus performante que les méthodes linéaires, non linéaires et de gradient. Elle donne des résultats proches des méthodes non linéaires utilisant les statistiques robustes et les résultats de calibration classique (F_{cal}). Ces dernières servent de résultats de comparaison pour les autres méthodes, puisque la calibration classique est considérée comme représentant la réalité physique inhérente à la scène.

5.1.2.2 Facteur de qualité

Bien que la méthode de comparaison présentée plus haut soit une très bonne méthode, elle ne peut suffir à nous renseigner sur l'efficacité d'une méthode par rapport à d'autres. Il faut donc comparer les résultats obtenus par différentes techniques. Une des techniques utilisées est appelée facteur de qualité, noté Q_F (Boufama et Mohr, 1995). Ici l'objectif est de mesurer la distance moyenne entre les points appariés qui ont été utilisés pour calculer la matrice fondamentale et les lignes épipolaires correspondantes. La mesure est donnée par

$$Q_F = \frac{\sum_{i=1}^n d(m'_i, Fm_i) + d(m_i, F^T m'_i)}{2n},$$

avec m'_i et m_i , $i = 1, \dots, n$, les points appariés dans les deux images de la paire stéréoscopique, et F la matrice fondamentale calculée.

Tableau 5.2 : Facteur de qualité pour différentes techniques de calcul de la matrice fondamentale

Q_f	Mire	Bureau	Bat	Lab
FlinN	1.70	2.16	1.98	1.13
Fh	0.99	1.02	1.23	0.82
Fmlin	1.33	2.08	1.90	0.96
Fllms	1.11	1.56	1.42	0.89

Les résultats sont donnés par le tableau 5.2. Nous constatons que la méthode

que nous avons développée donne de meilleurs résultats que les autres méthodes. Ceci vient du fait qu'on a utilisé explicitement les points appariés pour retrouver les lignes épipolaires par homographie (voir chapitre 2). La méthode que nous proposons permet donc d'avoir des lignes épipolaires qui sont très proches des points d'intérêt sélectionnés, ce qui présente un intérêt particulier pour l'application en synthèse de vues: l'intersection des lignes épipolaires lors de la reprojection des points dans la troisième vue se fait d'une manière plus précise (très proche de la position réelle du point d'intérêt qu'on veut projeter dans la nouvelle vue).

5.1.2.3 Temps de calcul

Un des principaux avantages des méthodes linéaires est le temps de calcul qui est souvent très faible comparativement au temps de calcul que nécessitent les méthodes non linéaires. Pour la paire d'images "Mire" (figure 5.1), nous avons établi un appariement entre 102 points et nous avons mesuré le temps de calcul pour les méthodes : linéaire, linéaire basée sur l'homographie et la méthode de moindre médiane des carrés qui est une méthode non linéaire utilisant les techniques de statistiques robustes. Sur un SPARC 5, pour la méthode linéaire, le temps de calcul est d'environ 0.26 ms, pour la méthode linéaire basée sur l'homographie, le temps est d'environ 0.37 ms alors que pour la méthode de moindre médiane des carrés il est d'environ 2.56 s. On constate que le gain en temps de calcul est appréciable pour le calcul de la matrice fondamentale lorsque nous utilisons la méthode linéaire basée sur l'homo-

graphie présentée dans la section 2.5. Bien que le temps de calcul pour la méthode linéaire sont inférieur à celui de la méthode que nous avons développée, le résultat obtenu par cette dernière est meilleur que celui par la méthode linéaire (la matrice fondamentale est plus proche de la matrice fondamentale réelle).

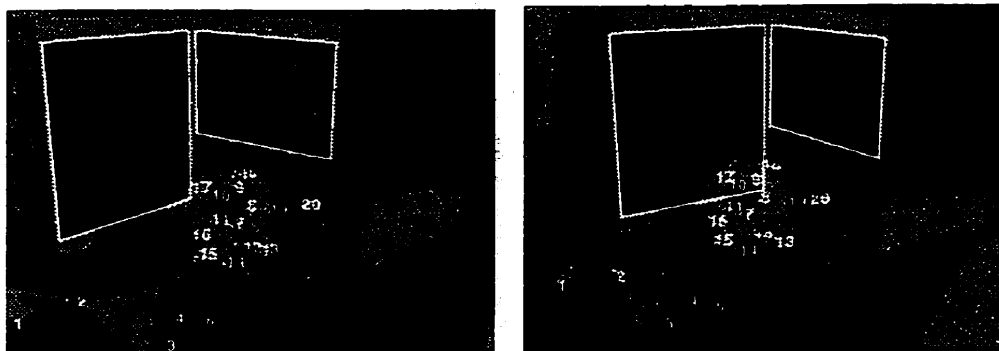
En conclusion la méthode que nous proposons présente des avantages sur la plupart des méthodes existantes, seule la méthode de moindre médiane des carrés donne des résultats meilleurs mais très proches lorsqu'on les compare par la technique de différence entre matrices fondamentales \mathbf{F}_{diff} . Toutefois les résultats obtenus par comparaison du facteur de qualité Q_F sont meilleurs dans le cas de la méthode linéaire basée sur l'homographie et le gain en temps de calcul est important.

5.2 Invariants projectifs

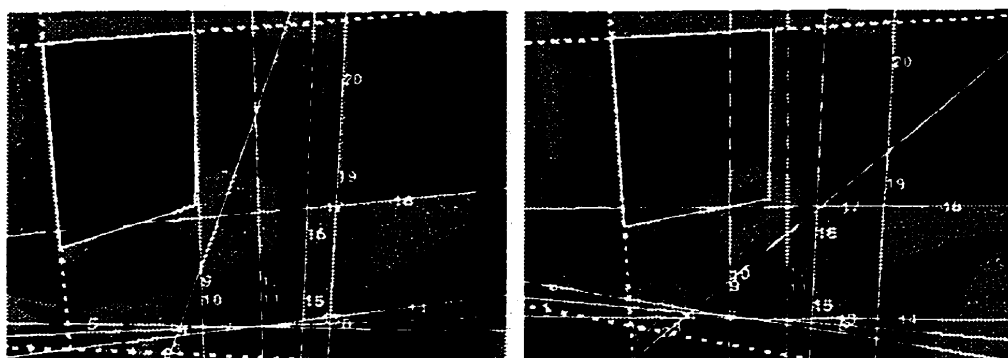
Dans cette section nous présentons les résultats obtenus par les méthodes de calcul des invariants projectifs.

Les images des figures 5.7 et 5.8 représentent les plans, points et lignes utilisés pour calculer les invariants projectifs par les méthodes présentées dans les sections 3.2 et 3.3. Les résultats sont donnés par les courbes des figures 5.9 et 5.10.

Pour calculer les invariants projectifs nous avons utilisé les paires d'images stéréoscopiques "Mire" et "Bureau". Pour le calcul de l'invariant basé sur un point de référence et deux homographies, nous avons sélectionné deux plans appariés sur les deux images d'une paire stéréoscopique (quadrilatères en blanc sur les figures 5.7(a)

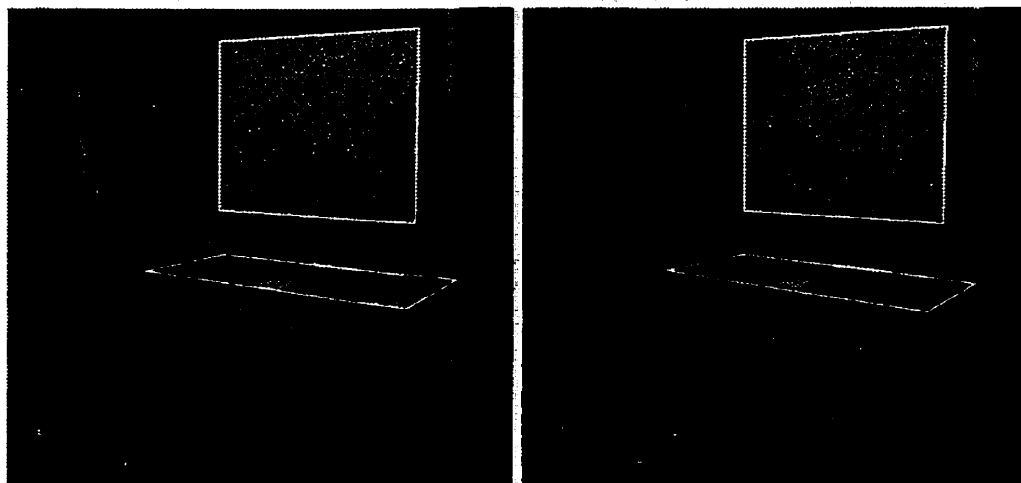


(a) Plans et points sélectionnés pour l'invariant basé sur un point et deux homographies

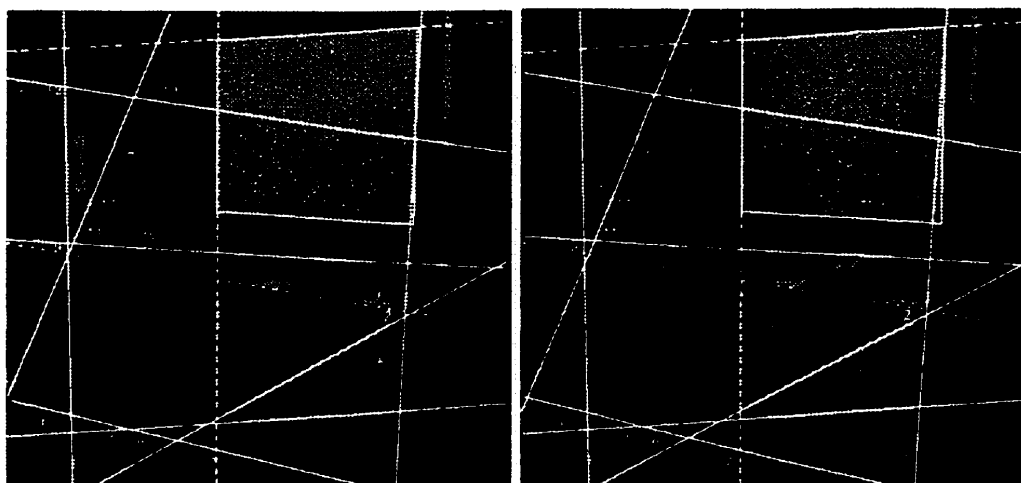


(b) Plan, points et lignes sélectionnés pour l'invariant basé sur une ligne et une homographie

Figure 5.7 : Paire stéréoscopique "Mire" utilisée pour le calcul des invariants projectifs

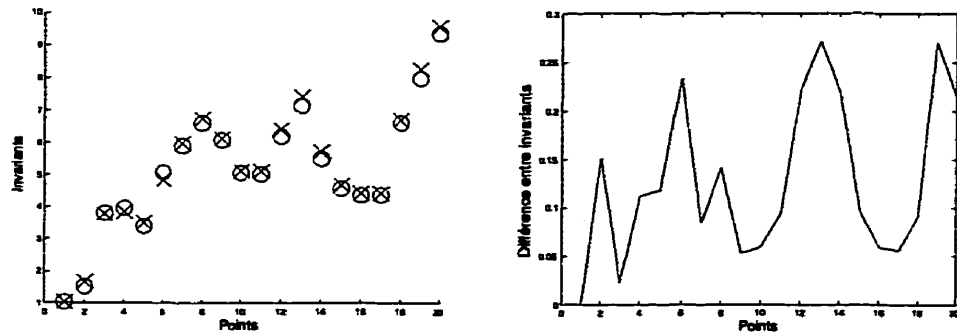


(a) Plans et points sélectionnés pour l'invariant basé sur un point et deux homographies

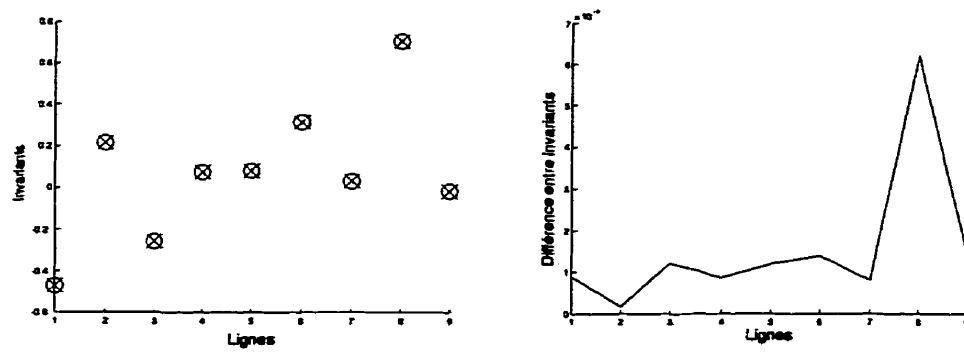


(b) Plan, points et lignes sélectionnés pour l'invariant basé sur une ligne et une homographie

Figure 5.8 : Paire stéréoscopique "Bureau" utilisée pour le calcul des invariants projectifs

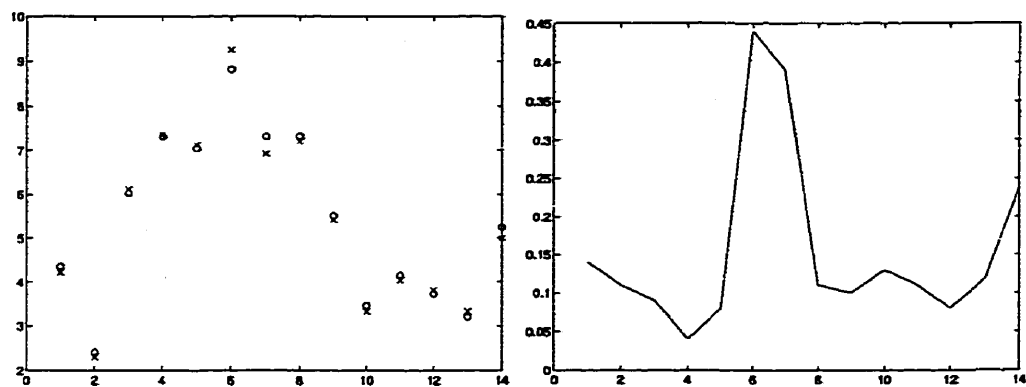


(a) Invariants basés sur un point de référence et deux homographies

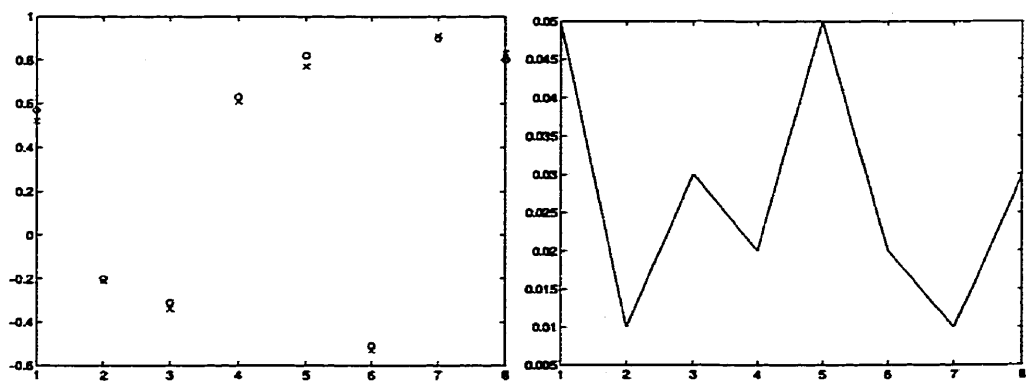


(b) Invariants basés sur une ligne de référence et une homographie

Figure 5.9 : Résultats de calcul des invariants projectifs pour la paire "Mire"



(a) Invariants basés sur un point de référence et deux homographies



(b) Invariants basés sur une ligne de référence et une homographie

Figure 5.10 : Résultats de calcul des invariants projectifs pour la paire "Bureau"

et 5.8 (a)) et un point de référence et son correspondant (le point dénoté par le symbole "+" sur les images des deux figures. Nous avons aussi sélectionné un ensemble de points appariés sur les images de la paire stéréoscopique, pour lesquels nous avons calculé l'invariant projectif. Les résultats sont donnés par les courbes des figures 5.9(a) et 5.10(a). La première courbe représente les valeurs de calcul des invariants obtenus pour les points appariés: "o" représente l'invariant pour les points de l'image gauche et "x" représente l'invariant pour les points de l'image droite. Nous constatons que les invariants obtenus pour les points appariés sont très proches, nous obtenons donc une même valeur invariante pour une paire de points appariés.

Une mesure qualitative de comparaison entre les invariants projectifs est la différence entre les valeurs obtenues pour chaque paire de points (ou lignes) appariés (Shashua, 1994a) :

$$\Delta = | I_l - I_r |,$$

où I_l c'est l'invariant calculé sur l'image gauche et I_r l'invariant calculé sur l'image droite.

Le résultat de comparaison par différence entre invariants pour les paires de points appariés dans le cas de l'invariant basé sur un point de référence et une homographie est donné par la deuxième courbe des figures 5.9(a) et 5.10(a). Nous constatons que le résultat est satisfaisant et que la précision obtenue est de l'ordre de 10^{-1} pixels. Ceci signifie que le processus d'appariement a été assez précis pour obtenir une très

bonne estimation des invariants projectifs.

Pour l'invariant basé sur une ligne de référence et une seule homographie nous avons sélectionné un plan apparié sur une paire d'images stéréoscopiques (quadrilatère en blanc sur les figures 5.7(b) et 5.8(b)), ce plan contient deux lignes (lignes pointillées sur les deux images des figures 5.7(b) et 5.8(b)). Ces deux lignes seront utilisées pour le calcul de l'invariant projectif comme décrit dans la section 3.3. Nous avons choisi la ligne connectant les points notés 1 et 2 comme étant la ligne de référence (ligne pointillée sur la table dans les images de la figure 5.7(b) et sur le bureau dans les images de la figure 5.8(b)). Nous avons sélectionné un ensemble de lignes appariées sur la paire d'images stéréoscopiques, pour lesquelles nous avons calculé l'invariant basé sur une ligne de référence et une seule homographie. Les résultats sont donnés par la première courbe des figures 5.9(b) et 5.10(b). Les valeurs obtenues pour les invariants projectifs sont très proches ("o" représente l'invariant pour les lignes de l'image gauche et "x" représente l'invariant pour les lignes de l'image droite). Nous constatons que la comparaison par la différence entre invariants, pour les lignes appariées, estimés en se basant sur une ligne de référence et une seule homographie, montre que les résultats obtenus sont satisfaisants (la deuxième courbe des figures 5.9(b) et 5.10(b)). Ceci montre que l'estimation de l'invariant projectif a donné un très bon résultat, ce qui montre que le processus d'appariement a été assez précis.

5.3 Synthèse de vues

Dans cette section nous présentons les résultats de l'application des techniques précédentes pour la synthèse de vues. Nous utilisons les méthodes décrites au chapitre 4 pour produire une nouvelle vue qui soit la plus proche possible de celle qui aurait été prise par une caméra placée au point de vue choisi. Le point de vue correspondant à la caméra virtuelle est défini par le choix d'une matrice de rotation \mathbf{R} et d'un vecteur de translation \mathbf{t} par rapport à la caméra de référence de la paire stéréoscopique. Les paramètres intrinsèques définis par la matrice \mathbf{A} sont ceux estimés préalablement par calibration (cas semi-calibré). Ceci permet de déterminer les matrices fondamentales \mathbf{F}_{iv} , $i = 1, 2$, entre la caméra virtuelle et les caméras de la paire stéréoscopique. Nous comparons par la suite les vues obtenues avec les vues réelles prises par des caméras placées aux endroits sélectionnés.

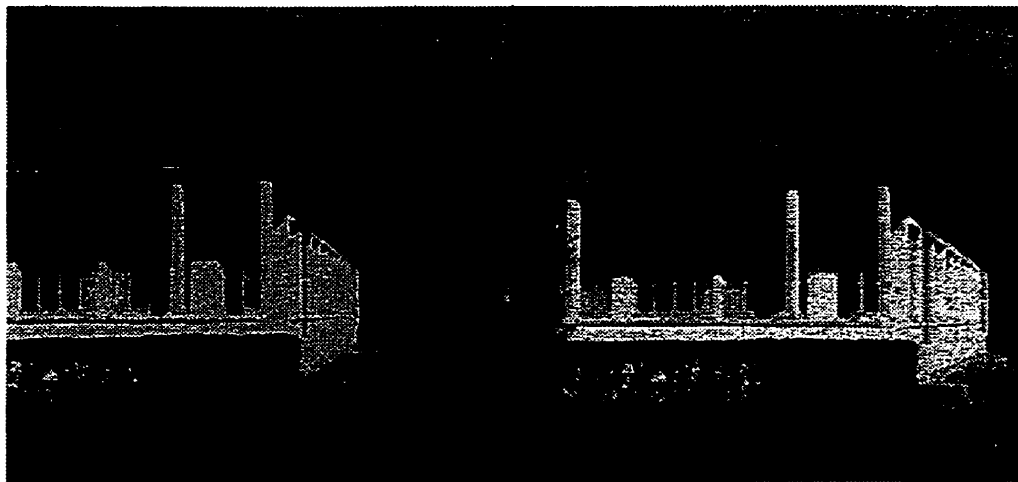


Figure 5.11 : Images sources pour la paire d'images "Bâtiment"

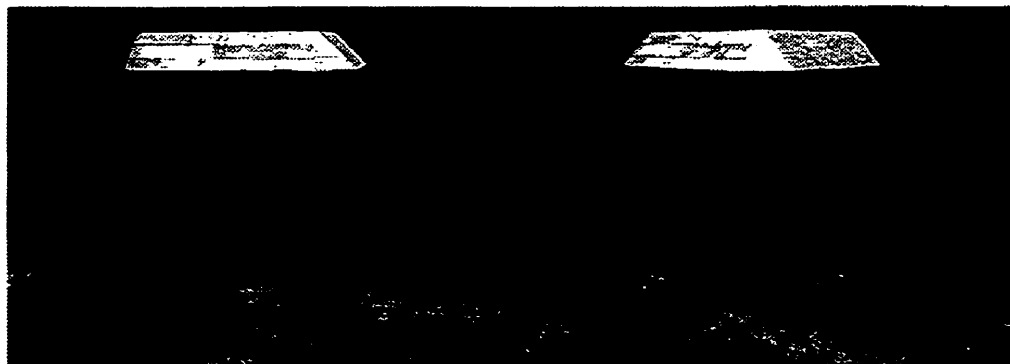


Figure 5.12 : Images sources pour la paire d'images "Maison"

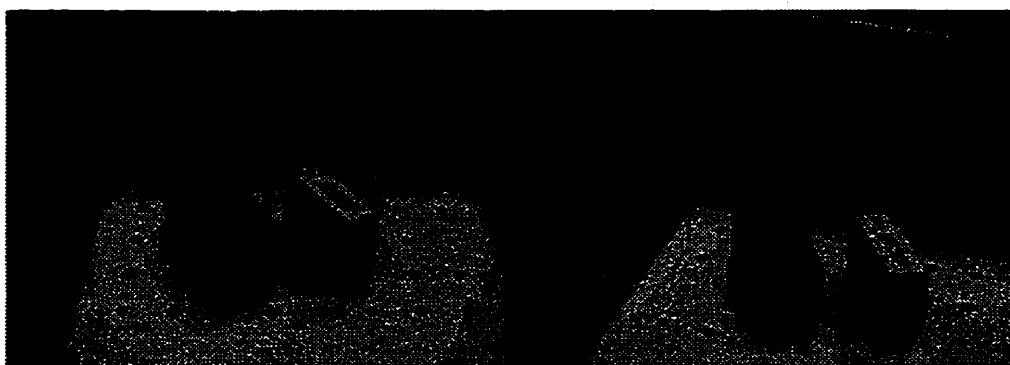


Figure 5.13 : Images sources pour la paire d'images "Lab"

Les figures 5.11, 5.12 et 5.13 montrent les images sources utilisées et les figures 5.14, 5.15 et 5.16 montrent les images résultantes en utilisant l'algorithme de synthèse de vues et les images réelles correspondantes. On constate que les images synthétisées sont visuellement proches des images réelles correspondantes.

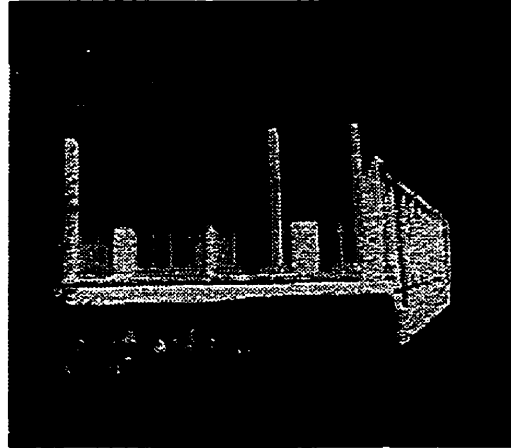
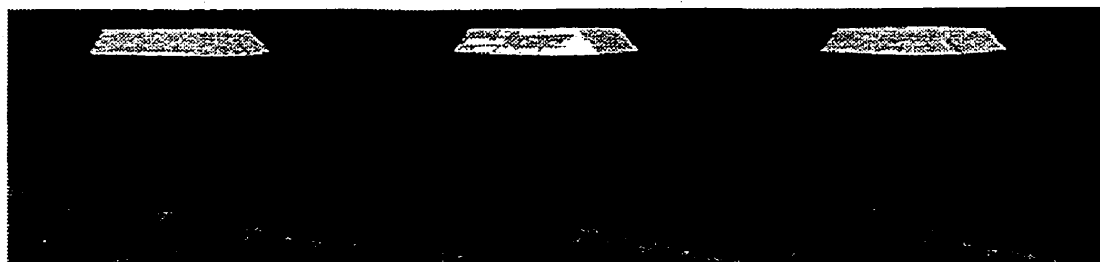


Figure 5.14 : Image synthétisée - "Bâtiment"

Pour comparer entre les vues synthétisées et les vues réelles correspondantes d'une manière qualitative nous avons utilisé la mesure suivante :

$$Q = \frac{\sum_{i=1}^n d(m_i, m'_i)}{n},$$

avec m_i le point d'intérêt sur l'image réelle, m'_i le point d'intérêt sur l'image synthétisée et n le nombre de points sélectionnés. La mesure Q représente l'erreur moyenne en pixels entre la position réelle du point image et sa position obtenue par reprojection. Pour l'image "Maison" (figure 5.15), cette erreur est approximative-



(a)

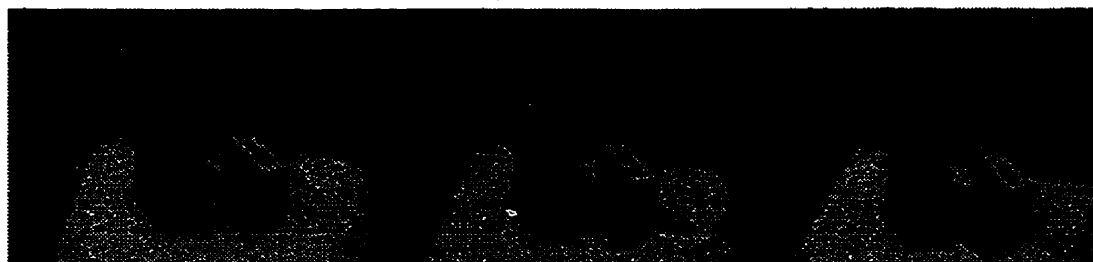


(b)

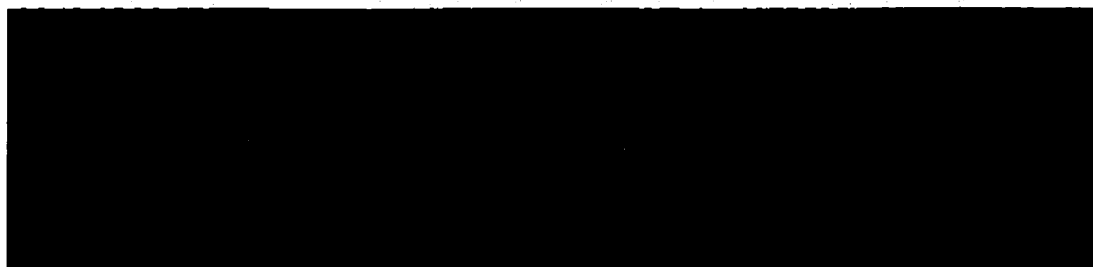


(c)

Figure 5.15 : Images synthétisées: (a) les images réelles; (b) images synthétisées par géométrie épipolaire; (c) images synthétisées par invariants projectifs - "Maison"



(a)



(b)



(c)

Figure 5.16 : Images synthétisées: (a) les images réelles; (b) images synthétisées par géométrie épipolaire; (c) images synthétisées par invariants projectifs - "Lab"

ment égale à 1.8 pixels pour la reprojection par la géométrie épipolaire et 2.0 pixels pour la reprojection par invariants projectifs. Pour l'image "Lab" (figure 5.16), cette erreur est d'environ 2.0 pixels pour la reprojection par géométrie épipolaire et d'environ 2.4 pixels pour la reprojection par invariants projectifs. On constate ainsi que l'application de la géométrie épipolaire donne de meilleurs résultats pour le transfert de points d'intérêt appariés sur une troisième vue. Toutefois, comme nous l'avons souligné au chapitre 4 la synthèse de nouvelles vues dans la région proche du plan trifocal est très complexe par les techniques de géométrie épipolaire, d'où l'intérêt d'utiliser les invariants projectifs pour la synthèse de nouvelles vues.

5.4 Conclusion

Dans ce chapitre nous avons présenté les résultats expérimentaux obtenus par les méthodes que nous avons développées dans les chapitres précédents. D'abord les différents résultats de comparaison montrent que la méthode de calcul de la matrice fondamentale décrite dans le chapitre 2 représente une très bonne alternative aux méthodes existantes de calcul de la matrice fondamentale surtout lorsque le temps de calcul est important. Nous avons présenté aussi les résultats pour les méthodes de calcul des invariants projectifs où nous montrons que les méthodes développées au chapitre 3 donnent une très bonne estimation de ces invariants. Enfin, nous avons montré que l'application des méthodes précédentes avec les techniques de *texturage bidimensionnel* en perspective permet de générer des nouvelles vues qui sont très

proches des vues réelles correspondantes.

Chapitre 6

Conclusion

Dans ce travail nous nous sommes intéressé à un domaine très récent de la vision artificielle, le rendu basé sur images. Il consiste en la synthèse de nouvelles vues les plus réalistes possibles à partir de vues existantes. Ici l'étude a porté sur le cas de génération de vues à partir des images prises par une paire stéréoscopique non étalonnée. Pour cela il faut comprendre les relations géométriques caractérisant une configuration de caméras stéréoscopiques.

La géométrie projective nous a permis d'avoir une description des relations entre caméras par le biais de la géométrie épipolaire et des invariants projectifs. Les outils de la géométrie projective permettent une description *valide* de la géométrie de la scène, et peuvent être utilisés pour une synthèse de nouvelles vues satisfaisant les contraintes géométriques inhérentes à la configuration des caméras et de la scène. La vue ainsi obtenue est plus proche de la vue qui serait capturée par une caméra réelle

placée à l'endroit choisi pour placer la caméra virtuelle. L'étude de ce sujet nous a amené à développer des outils permettant la génération de nouvelles vues d'un réalisme satisfaisant.

Dans ce sens, notre contribution se situe principalement à trois niveaux (la géométrie épipolaire, l'invariance projective et la synthèse de nouvelles vues à partir d'un appariement épars entre les images d'une paire stéréoscopique) menant à la génération de nouvelles vues à partir d'images sources prises par une paire de caméras stéréoscopiques non calibrées en se basant plus particulièrement sur l'exploitation des propriétés de la scène. Ce premier élément consiste en l'utilisation de la contrainte de planarité associée à la présence de surfaces planes dans la scène. Cette contrainte est caractérisée par une relation homographique entre les projections sur les plans images d'une même surface plane. Cette transformation contient toute l'information nécessaire sur la perspective de la surface plane.

Tout d'abord, comme contribution au niveau de la géométrie épipolaire, nous avons utilisé la contrainte de planarité pour définir la géométrie épipolaire entre deux caméras d'une paire stéréoscopique caractérisée par la matrice fondamentale. La relation homographique entre les deux images d'un même plan présent dans deux plans images différents nous permet d'établir les relations épipolaires qui nous sont nécessaires pour la résolution de l'équation de contrainte épipolaire. Les homographies permettent de déterminer les lignes épipolaires caractérisant une configuration de caméras donnée, ainsi que les points épipolaires correspondants. Ceci amène par

la suite à la définition d'un paramétrage pour la matrice fondamentale en termes de points épipolaires. Ce paramétrage permet l'obtention d'une forme de la matrice fondamentale qui satisfait la contrainte de rang 2 caractéristique d'une matrice fondamentale (la forme de la matrice fondamentale fait en sorte que son rang soit égal à 2, cette propriété est la source la plus importante d'instabilité calculatoire). Une fois cette étape franchie, une résolution d'un système d'équations linéaires définies par les contraintes épipolaires permet de retrouver les coefficients de la matrice fondamentale d'une manière numériquement stable. La méthode proposée traite les différents types de configurations de caméras par l'exploitation des propriétés de la géométrie projective. Les résultats de comparaisons que nous avons présentés montrent bien que la méthode développée dans ce travail donne des résultats satisfaisants proches des résultats de calibration classique (considérés comme les valeurs réelles dans nos tests) et des techniques non linéaires utilisant les statistiques robustes (moindre médiane des carrés). En plus, elle a l'avantage de permettre un gain appréciable en temps de calcul.

Notre deuxième contribution se situe au niveau des invariants projectifs, où nous avons présenté deux techniques permettant le calcul des invariants à partir de l'appariement entre des primitives d'intérêt (points et lignes) présentes sur deux images d'une paire stéréoscopique. Ici aussi nous avons explicitement utilisé la contrainte de planarité pour définir ces invariants projectifs. D'abord nous avons montré comment nous avons obtenu un invariant pour les points en se basant sur un point de référence

et deux homographies caractérisant deux plans de la scène. Cet invariant est défini en terme de birapports, l'invariant de base en géométrie projective. Pour les lignes, nous avons développé une nouvelle méthode permettant de calculer un invariant projectif à partir d'un appariement entre lignes, en terme de birapports calculés en utilisant une ligne de référence et une seule homographie correspondant à un plan de la scène. Les résultats de comparaison de la différence entre invariants pour les primitives d'intérêt montrent que les techniques proposées donnent une très bonne estimation de ces invariants. Les résultats obtenus dans le chapitre 5 montrent que cette estimation est très précise.

Finalement, comme troisième contribution nous avons présenté une nouvelle approche pour la synthèse de nouvelles vues à partir de vues connues et ce sans modèle tridimensionnel. Cette approche est basée sur la décomposition de la scène en morceaux de surfaces planes. Elle a l'avantage de n'utiliser qu'un appariement éparé pour produire la nouvelle vue. Ici nous avons exploité les outils de géométrie projective développés plus haut, c'est-à-dire la matrice fondamentale et les invariants projectifs pour reprojeter les points appariés sur une nouvelle vue. L'appariement éparse permet de représenter l'image par un maillage dont les points appariés sont les points de contrôle. Ainsi, la reprojection des points appariés permet de définir un maillage sur la nouvelle vue qui sera rempli en utilisant des techniques de texturage bidimensionnel en perspective. Nous évitons ainsi d'avoir à faire un appariement dense entre les images de la paire stéréoscopique. La technique de texturage utilisée permet une bonne

reprojection lorsque la surface d'intérêt est plane, d'où l'intérêt d'une décomposition de la scène en morceaux de surfaces de ce type. Cette hypothèse est souvent satisfaite (scènes urbaines, scènes industrielles, etc.). Notre méthode peut être aussi utilisée dans le cas de scènes complexes où les zones d'intérêt contiennent souvent des objets qu'on peut décomposer en surfaces planes. La faisabilité de l'approche développée a été démontrée sur plusieurs exemples qui montrent que nous obtenons une bonne précision sur la position des points dans l'image synthétisée. D'une manière subjective nous pouvons visuellement constater que les vues générées sont très proches des vues réelles correspondantes.

Le travail que nous avons présenté soulève des questions qui peuvent être abordées dans de futures recherches, aussi bien au niveau des outils de la géométrie projective que de la synthèse de vues.

Au niveau de la géométrie épipolaire, l'intégration de techniques de statistiques robustes qui permettent la détection et ensuite le rejet de faux appariements améliorerait les résultats que nous avons obtenu par la méthode basée sur l'homographie. D'un autre côté l'utilisation de la géométrie épipolaire déterminée à partir d'un petit nombre de points permettrait d'établir un appariement des autres points de manière automatique et serait intéressant de le considérer dans un schéma itératif global dans lequel le passage de l'appariement à matrice fondamentale et de matrice fondamentale à l'appariement permettrait éventuellement une amélioration des résultats obtenus.

Du côté de la contrainte de planarité, il serait intéressant d'étudier la possibilité

de relâchement d'une telle contrainte pour traiter un cas plus général. Ceci est aussi valable pour les invariants projectifs qui exploitent cette même contrainte.

Pour la synthèse de vues un schéma d'interpolation plus général nous aiderait à traiter des scènes sans avoir à les décomposer en morceaux de surfaces planes. L'idée est de développer une technique de texturage bidimensionnel capable de reprojeter en perspective non seulement de plans mais aussi des formes plus complexes. Enfin, une question importante à considérer dans le futur est le développement d'un modèle de changement d'illumination lors de la synthèse de vues sans reconstruction tridimensionnelle explicite. Ainsi, la vue générée sera plus proche de la vraie vue qui serait prise par une caméra placée à l'endroit choisi: non seulement la position des points reprojétés sera précise, mais aussi l'interaction de la source de lumière avec la scène sera modélisée (couleur, ombres, etc.), et la vue obtenue sera plus réaliste. Malgré que les modèles d'illumination ont été très étudiés en infographie, l'application de la géométrie projective pourrait éventuellement aider à définir un modèle à partir seulement des images bidimensionnelles et ce sans reconstruction tridimensionnelle explicite.

Bibliographie

- ADELSON, E. (1995), Layered representations for vision and video, dans *Proceedings of the IEEE Workshop on Representation of Visual Scenes*, Cambridge, MA, USA.
- ADIV, G. (1985), Determining three-dimensional motion and structure from optical flow generated by several moving objects, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 7(4), 384–400.
- AKHLOUFI, M., COHEN, P. et POLOTSKI, V. (1999), Novel views from non-calibrated stereo, dans *Proceedings of the Vision Interface*, Trois Rivières, QC, Canada. Accepted for publication.
- AKHLOUFI, M., POLOTSKI, V. et COHEN, P. (1999), Virtual view synthesis from uncalibrated stereo cameras, dans *Proceedings of the IEEE International Conference on Multimedia, Computing and Systems*, Firenze, Italy. Accepted for publication.

- AKHLOUFI, M., TONG, W., POLOTSKI, V. et COHEN, P. (1998), Estimating fundamental matrix based on planar objects in stereo, dans *Proceedings of the Joint Conference on Information Sciences, First International Workshop on Computer Vision, Pattern Recognition and Image Processing*, Durham, NC, USA, pp. 398–400.
- AVIDAN, S., EVGENIOU, T., SHASHUA, A. et POGGIO, T. (1997), Image-based view synthesis by combining trilinear tensors and learning techniques, dans *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, Lausanne, Switzerland, pp. 103–10.
- AVIDAN, S. et SHASHUA, A. (1997), Novel view synthesis in tensor space, dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp. 1034–1040.
- BAKER, S., SZELISKI, R. et ANANDAN, P. (1998), A layered approach to stereo reconstruction, dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA, USA, pp. 434–441.
- BERGER, M., SIMON, G., PETITJEAN, S. et WROBEL-DAUTCOURT, B. (1996), Mixing synthesis and video images of outdoor environments: Application to the bridges of paris, dans *Proceedings of the International Conference on Pattern Recognition*, Vienna, Austria, p. A7M.6.

- BOUFAMA, B. et MOHR, R. (1995), Epipole and fundamental matrix estimation using the virtual parallax property, dans *Proceedings of the IEEE International Conference on Computer Vision*, Boston, MA, USA, pp. 1030–1036.
- BOUGNOUX, S. (1998), From projective to euclidean space under any practical situation, a criticism of self-calibration, dans *Proceedings of the IEEE International Conference on Computer Vision*, Bombay, India, pp. 790–795.
- CARLSSON, S. (1995), Duality of reconstruction and positioning from projective views, dans *Proceedings of the IEEE Workshop on Representation of Visual Scenes*, Cambridge, MA, USA.
- CHANG, N. et ZAKHOR, A. (1995), Arbitrary view generation for three-dimensional scenes from uncalibrated video cameras, dans *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Detroit, MI, USA, pp. 2455–2458.
- CHANG, N. et ZAKHOR, A. (1997), View generation for three-dimensional scenes from video sequences, *IEEE Transactions on Image Processing* 6(4), 584–598.
- CHEN, S. et WILLIAMS, L. (1993), View interpolation for image synthesis, dans *Computer Graphics Proceedings*, Anaheim, CA, USA, pp. 279–288.
- COHEN, P., HERVÉ, J.-Y. et AKHLOUFI, M. (1998), Augmented reality concepts for mining vehicle operation, dans *Proceedings of the CIM/CMMI/MIGA Conference*, Montréal, QC, Canada, p. N088.

- CSURKA, G. et FAUGERAS, O. (1998), Computing 3-dimensional project invariants from a pair of images using the grassmann-cayley algebra, *Image and Vision Computing* **16**(1), 3–12.
- CSURKA, G., ZELLER, C., ZHANG, Z. et FAUGERAS, O. (1997), Characterizing the uncertainty of the fundamental matrix, *Computer Vision and Image Understanding* **68**(1), 18–36.
- DEBEVEC, P.E. TAYLOR, C. et MALIK, J. (1996), Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach, dans *Computer Graphics Proceedings*, New Orleans, LA, USA, pp. 11–20.
- DERICHE, R., ZHANG, Z., LUONG, Q. et FAUGERAS, O. (1994), Robust recovery of the epipolar geometry for an uncalibrated stereo rig, dans *Proceedings of the European Conference on Computer Vision*, Stockholm, Sweden, pp. A:567–576.
- DYER, C. (1997), Image-based scene rendering and manipulation research at the university of wisconsin, dans *Proceedings of the DARPA Image Understanding Workshop*, New Orleans, LA, USA, pp. 63–67.
- FAUGERAS, O. (1992), What can be seen in three dimensions with an uncalibrated stereo rig?, dans *Proceedings of the European Conference on Computer Vision*, Santa Margherita Ligure, Italy, pp. 563–578.
- FAUGERAS, O. (1993), *Three-Dimensional Computer Vision*, MIT Press, Cambridge, MA, USA.

- FAUGERAS, O., LUONG, Q. et MAYBANK, S. (1992), Camera self-calibration: Theory and experiments, dans *Proceedings of the European Conference on Computer Vision*, Santa Margherita Ligure, Italy, pp. 321-334.
- FAUGERAS, O. et ROBERT, L. (1996), What can two images tell us about a third one?, *International Journal of Computer Vision* 18(1), 5-19.
- FAUGERAS, O., ROBERT, L., LAVEAU, S., CSURKA, G., ZELLER, C., GAUCLIN, C. et ZOGHLAMI, I. (1998), 3-d reconstruction of urban scenes from image sequences, *Computer Vision and Image Understanding* 69(3), 292-309.
- FOLEY, J., VAN DAM, A., FEINER, S. et HUGHES, J. (1990), *Computer Graphics: Principles and Practice*, 2nd edn, Addison-Wesley, Reading, MA, USA.
- FORSYTH, D., MUNDY, J., ZISSERMAN, A. et BROWN, C. (1990), Invariance: A new framework for vision, dans *Proceedings of the IEEE International Conference on Computer Vision*, Osaka, Japan, pp. 598-605.
- GOLUB, G. et VAN LOAN, C. (1989), *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, USA.
- HARTLEY, R. (1992a), Calibration of camera using the essential matrix, dans *Proceedings of the DARPA Image Understanding Workshop*, San Diego, CA, USA, pp. 911-915.

- HARTLEY, R. (1992*b*), Estimation of relative camera positions for uncalibrated camera, dans *Proceedings of the European Conference on Computer Vision*, Santa Margherita Ligure, Italy, pp. 579–587.
- HARTLEY, R. (1993), Camera calibration using line correspondences, dans *Proceedings of the DARPA Image Understanding Workshop*, Washington, DC, USA, pp. 361–366.
- HARTLEY, R. (1994*a*), Euclidean reconstruction from uncalibrated views, dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, pp. 908–912.
- HARTLEY, R. (1994*b*), Projective reconstruction and invariants from multiple images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**(10), 1036–1041.
- HARTLEY, R. (1994*c*), Projective reconstruction from line correspondences, dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, pp. 903–907.
- HARTLEY, R. (1995), A linear method for reconstruction from lines and points, dans *Proceedings of the IEEE International Conference on Computer Vision*, Cambridge, MA, USA, pp. 882–887.
- HARTLEY, R. (1997*a*), In defense of the eight-point algorithm, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(6), 580–593.

- HARTLEY, R. (1997b), Kruppa's equations derived from the fundamental matrix, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(2), 133–135.
- HARTLEY, R. (1997c), Lines and points in three views and the trifocal tensor, *International Journal of Computer Vision* **22**(2), 125–140.
- HARTLEY, R. (1997d), Self-calibration of stationary cameras, *International Journal of Computer Vision* **22**(1), 5–23.
- HECKBERT, P. (1986), Survey of texture mapping, *IEEE Computer Graphics and Applications* **6**(11), 56–67.
- HLAVAC, V., LEONARDIS, A. et WERNER, T. (1996), Automatic selection of reference views for image-based scene representations, dans *Proceedings of the European Conference on Computer Vision*, Cambridge, UK, pp. I:516–535.
- HORAUD, R. et CSURKA, G. (1998), Self-calibration and euclidean reconstruction using motions of a stereo rig, dans *Proceedings of the IEEE International Conference on Computer Vision*, Bombay, India, pp. 96–103.
- HORN, B. (1986), *Robot Vision*, MIT Press, Cambridge, MA, USA.
- HUANG, T. et NETRAVALI, A. (1994), Motion and structure from feature correspondences: A review, *Proceedings of the IEEE* **82**, 252–268.

- IRANI, M., ANANDAN, P. et HSU, S. (1995), Mosaic based representations of video sequences and their applications, dans *Proceedings of the IEEE International Conference on Computer Vision*, Cambridge, MA, USA, pp. 605-611.
- JAIN, R. et WAKIMOTO, K. (1995), Multiple perspective interactive video, dans *Proceedings of the IEEE Conference on Multimedia Systems*, Washington, DC, USA.
- KANADE, T. (1995), Development of a video-rate stereo machine, dans *Proceedings of the IEEE Conference on Intelligent Robots and Systems*, Pittsburgh, PA, USA, pp. 95-100.
- KANADE, T. (1996), Immersion into visual media: new applications of image understanding, *IEEE Expert* 11(1), 73-80.
- KANADE, T., NARAYANAN, I. et RANDER, P. (1996), Virtualized reality: being mobile in a visual scene, dans J. Ponce, A. Zisserman et M. Hebert, éditeurs, *Object Representation in Computer Vision II. European Conference on Computer Vision International Workshop Proceedings*, Cambridge, UK, pp. 273-85.
- KANADE, T., NARAYANAN, P. et RANDER, P. (1995), Virtualized reality: concepts and early results, dans *Proceedings of the IEEE Workshop on Representation of Visual Scenes*, Cambridge, MA, USA, pp. 69-76.
- KANADE, T., RANDER, P. et NARAYANAN, P. (1997), Virtualized reality: constructing virtual worlds from real scenes, *IEEE Multimedia* 4(1), 34-47.

- KANADE, T., YOSHIDA, A., ODA, K., KANO, H. et TANAKA, M. (1996), A stereo machine for video-rate dense depth mapping and its new applications, dans *Proceedings of the IEEE Computer Vision and Pattern Recognition*, San Fransisco, CA, USA, pp. 196–202.
- KANG, S. (1997), A survey of image-based rendering techniques, rapport technique CRL 97/4, Digital Equipment Corporation, Cambridge Research Lab.
- KANG, S. et SZELISKI, R. (1997), 3-d scene data recovery using omnidirectional multibaseline stereo, *International Journal of Computer Vision* 25(2), 167–183.
- KATKERE, A., MOEZZI, S., KURARA, D., KELLY, P. et JAIN, R. (1997), Toward video-based immersive environments, *Multimedia Systems* 5(2), 69–85.
- KOCH, R. (1995), 3d surface reconstruction from stereoscopic image sequences, dans *Proceedings of the IEEE International Conference on Computer Vision*, Cambridge, MA, USA, pp. 109–114.
- KUMAR, R., ANANDAN, P., IRANI, M., BERGEN, J. et HANNA, K. (1995), Representation of scenes from collections of images, dans *Proceedings of the IEEE Workshop on Representation of Visual Scenes*, Cambridge, MA, USA.
- KUTULAKOS, K. (1995), Affine surface reconstruction by purposive viewpoint control, dans *Proceedings of the IEEE International Conference on Computer Vision*, Cambridge, MA, USA, pp. 894–901.

- LAVEAU, S. et FAUGERAS, O. (1994), 3-d scene representation as a collection of images, dans *Proceedings of the International Conference on Pattern Recognition*, Jerusalem, Israel, pp. A:689–691.
- LAVEAU, S. et FAUGERAS, O. (1996), Oriented projective geometry for computer vision, dans *Proceedings of the European Conference on Computer Vision*, Cambridge, UK, pp. I:147–156.
- LE BRAS-MEHLMAN, E., SCHMITT, M., BOISSONNAT, J. et FAUGERAS, O. (1988), How the delaunay triangulation can be used for representing stereo data, dans *Proceedings of the International Conference on Computer Vision*, Tarpon Springs, FL, USA, pp. 54–63.
- LENGYEL, J. (1998), The convergence of graphics and vision, *IEEE Computer* **31**(7), 46–53.
- LEVOY, M. et HANRAHAN, P. (1996), Light field rendering, dans *Computer Graphics Proceedings*, pp. 31–42.
- LONGUET-HIGGINS, H. (1981), A computer algorithm for reconstructing a scene from two projections, *Nature* **293**, 133–135.
- LUONG, Q. et FAUGERAS, O. (1992), Self-calibration of a camera using multiple images, dans *Proceedings of the International Conference on Pattern Recognition*, The Hague, Netherlands, pp. I:9–12.

- LUONG, Q. et FAUGERAS, O. (1994a), On the direct determination of epipoles: A case study in algebraic methods for geometric problems, dans *Proceedings of the International Conference on Pattern Recognition*, Jerusalem, Israel, pp. A:243–247.
- LUONG, Q. et FAUGERAS, O. (1994b), A stability analysis of the fundamental matrix, dans *Proceedings of the European Conference on Computer Vision*, Stockholm, Sweden, pp. A:577–588.
- LUONG, Q. et FAUGERAS, O. (1996), The fundamental matrix: Theory, algorithms, and stability analysis, *International Journal of Computer Vision* 17(1), 43–75.
- LUONG, Q. et VIEVILLE, T. (1994), Canonic representations for the geometries of multiple projective views, dans *Proceedings of the European Conference on Computer Vision*, Stockholm, Sweden, pp. A:589–599.
- MASE, K. (1996), Computing field-of-view of stitched panorama to create fov sensitive virtual environments, dans *Proceedings of the International Conference on Pattern Recognition*, Vienna, Austria, p. A72.3.
- MAYBANK, S. (1995), Relation between 3d invariants and 2d invariants, dans *Proceedings of the IEEE Workshop on Representation of Visual Scenes*, Cambridge, MA, USA.

- MCMILLAN, L. et BISHOP, G. (1995), Plenoptic modeling: an image-based rendering system, dans *Computer Graphics Proceedings*, Los Angeles, CA, USA, pp. 39–46.
- MITICHE, M. (1994), *Computational Analysis of Visual Motion*, Plenum Press, New York, NY, USA.
- MOEZZI, S., KATKERE, A., KURARA, D. et JAIN, R. (1996a), An emerging medium: interactive three-dimensional digital video, dans *Proceedings of the IEEE International Conference on Multimedia, Computing and Systems*, Hiroshima, Japan, pp. 358–361.
- MOEZZI, S., KATKERE, A., KURARA, D. et JAIN, R. (1996b), Immersive video, dans *Proceedings of the IEEE Virtual Reality Annual International Symposium*, Santa Clara, CA, USA, pp. 17–24.
- MOEZZI, S., KATKERE, A., KURARA, D. et JAIN, R. (1996c), Reality modeling and visualization from multiple video sequences, *IEEE Computer Graphics and Applications* 16(10), 58–63.
- MOEZZI, S., LI-CHENG, T. et GERARD, P. (1997), Virtual view generation for 3d digital video, *IEEE Multimedia* 4, 18–26.
- MUNDY, J., ZISSERMAN, A. et FORSYTH, D. (1993), *Applications of Invariance in Computer Vision*, Springer-Verlag, Berlin, Germany.

- NARAYANAN, P. et KANADE, T. (1998), Virtual worlds using computer vision, dans *Proceedings of the IEEE and ATR Workshop on Computer Vision for Virtual Reality Based Human Communication*, Cambridge, MA, USA, pp. 2-13.
- NARAYANAN, P., RANDER, P. et KANADE, T. (1998), Constructing virtual worlds using dense stereo, dans *Proceedings of the IEEE International Conference on Computer Vision*, Bombay, India, pp. 3-10.
- NIEM, W. et BROSZIO, H. (1995), Mapping texture from multiple camera views onto 3d-object models for computer animation, dans *Proceedings of the International Workshop on Stereoscopic and Three Dimensional Imaging*, Santorini, Grèce.
- OKUTOMI, M. et KANADE, T. (1993), A multiple-baseline stereo, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **15**(4), 353-363.
- PELEG, S. et HERMAN, J. (1997), Panoramic mosaics by manifold projection, dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp. 338-343.
- RANDER, P., NARAYANAN, P. et KANADE, T. (1997), Virtualized reality: constructing time-varying virtual worlds from real world events, dans R. Yagel et H. Hagen, éditeurs, *Proceedings of the IEEE Visualization*, Phoenix, AZ, USA, pp. 277-84.

- RANDER, P., NARAYANAN, P. et KANADE, T. (1998), Constructing virtual worlds using dense stereo, dans *Proceedings of the IEEE International Conference on Computer Vision*, Bombay, India, pp. 3–10.
- ROUSSEEUW, P. et LEROY, A. (1987), *Robust Regression and Outliers Detection*, John Wiley, New York, NY, USA.
- ROUSSO, B., PELEG, S., FINCI, I. et RAV-ACHA, A. (1998), Universal mosaicing using pipe projection, dans *Proceedings of the IEEE International Conference on Computer Vision*, Bombay, India, pp. 945–952.
- SAWHNEY, H. et KUMAR, R. (1997), True multi-image alignment and its application to mosaicing and lens distortion correction, dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp. 450–456.
- SCHARSTEIN, D. (1996), Stereo vision for view synthesis, dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, pp. 852–858.
- SEITZ, S. (1997), Image-based transformation of viewpoint and scene appearance, Thèse de doctorat, Computer Science Dept, University of Wisconsin, Madison, WI, USA. <http://www.cs.wisc.edu/computer-vision/pubs.html>.

- SEITZ, S. et DYER, C. (1995a), Complete scene structure from four point correspondences, dans *Proceedings of the IEEE International Conference on Computer Vision*, Cambridge, MA, USA, pp. 330–337.
- SEITZ, S. et DYER, C. (1995b), Physically-valid view synthesis by image interpolation, dans *Proceedings of the IEEE Workshop on Representation of Visual Scenes*, Cambridge, MA, USA, pp. 18–25.
- SEITZ, S. et DYER, C. (1996a), Toward image-based scene representation using view morphing, dans *Proceedings of the International Conference on Pattern Recognition*, Vienna, Austria, pp. I: A 84–89.
- SEITZ, S. et DYER, C. (1996b), View morphing, dans *Computer Graphics Proceedings*, New Orleans, LA, USA, pp. 21–30.
- SEITZ, S. et DYER, C. (1997a), Photorealistic scene reconstruction by voxel coloring, dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp. 1067–1073.
- SEITZ, S. et DYER, C. (1997b), View morphing: uniquely predicting scene appearance from basis images, dans *Proceedings of the DARPA Image Understanding Workshop*, New Orleans, LA, USA, pp. 881–887.
- SEITZ, S. et KUTULAKOS, K. (1998), Plenoptic image editing, dans *Proceedings of the IEEE International Conference on Computer Vision*, Bombay, India, pp. 17–24.

SEMPLE, J. et KNEEBONE, G. (1979), *Algebraic Projective Geometry*, Oxford Clarendon Press, Oxford, UK.

SHASHUA, A. (1993), Projective depth: A geometric invariant for 3d reconstruction from two perspective/orthographic views for visual recognition, dans *Proceedings of the IEEE International Conference on Computer Vision*, Berlin, Germany, pp. 583–590.

SHASHUA, A. (1994a), Projective structure from uncalibrated images: Structure-from-motion and recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16(8), 778–790.

SHASHUA, A. (1994b), Trilinearity in visual recognition by alignment, dans *Proceedings of the European Conference on Computer Vision*, Stockholm, Sweden, pp. A:479–484.

SHASHUA, A. (1995), Algebraic functions for recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17(8), 779–789.

SHASHUA, A. (1997), Trilinear tensor: the fundamental construct of multiple-view geometry and its applications, dans G. Sommer et J. Koenderink, éditeurs, *Algebraic Frames for the Perception-Action Cycle. International Workshop Proceedings*, Kiel, Germany, pp. 190–206.

- SHASHUA, A. et NAVAB, N. (1996), Relative affine structure: Canonical model for 3d from 2d geometry and applications, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**(9), 873–883.
- SHASHUA, A. et WERMAN, M. (1995), Trilinearity of three perspective views and its associated tensor, dans *Proceedings of the IEEE International Conference on Computer Vision*, Cambridge, MA, USA, pp. 920–925.
- SHUM, H., HAN, M. et SZELISKI, R. (1998), Interactive construction of 3d models from panoramic mosaics, dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA, USA, pp. 427–433.
- SHUM, H. et SZELISKI, R. (1998), Construction and refinement of panoramic mosaics with global and local alignment, dans *Proceedings of the IEEE International Conference on Computer Vision*, Bombay, India, pp. 953–958.
- SZELISKI, R. (1996), Video mosaics for virtual environments, *IEEE Computer Graphics and Applications* **16**(2), 22–30.
- SZELISKI, R. (1998), Building shape and appearance models from multiple images, dans *Proceedings of the ACM Workshop on image-based modeling and rendering*, Stanford University, CA, USA.
- SZELISKI, R. et SHUM, H.-Y. (1997), Creating full view panoramic image mosaics and environment maps, dans *Computer Graphics Proceedings*, Los Angeles, CA, USA, pp. 251–8.

- TAYLOR, C., DEBEVEC, P. et MALIK, J. (1996a), Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach, dans *Computer Graphics Proceedings*, New Orleans, LA, USA.
- TAYLOR, C., DEBEVEC, P. et MALIK, J. (1996b), Reconstructing polyhedral models of architectural scenes from photographs, dans *Proceedings of the European Conference on Computer Vision*, Cambridge, UK, pp. II:659–668.
- TONG, W., AKHLOUFI, M., POLOTSKI, V. et COHEN, P. (1998), Two useful invariants in computer vision and their applications, dans *Proceedings of the Joint Conference on Information Sciences, First International Workshop on Computer Vision, Pattern Recognition and Image Processing*, Durham, NC, USA, pp. 401–404.
- TORR, P. et MURRAY, D. (1997), The development and comparison of robust methods for estimating the fundamental matrix, *International Journal of Computer Vision* **24**(3), 271–300.
- TORR, P. et ZISSERMAN, A. (1997), Robust parameterization and computation of the trifocal tensor, *Image and Vision Computing* **15**(8), 591–605.
- TORR, P., ZISSERMAN, A. et MAYBANK, S. (1998), Robust detection of degenerate configurations while estimating the fundamental matrix, *Computer Vision and Image Understanding* **71**(3), 312–333.

- TORR, P. et ZISSERMANN, A. (1997), Performance characterization of fundamental matrix estimation under image degradation, *Machine Vision and Applications* 9(5-6), 321-333.
- TSAI, R. (1987), A versatile camera calibration technique for high-accuracy 3d machine vision metrology using of-the-shelf tv cameras and lenses, *IEEE Journal of Robotics and Automation* 3(4), 323-344.
- ULLMAN, S. et BASRI, R. (1991), Recognition by linear combinations of models, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(10), 992-1006.
- VIEVILLE, T., FAUGERAS, O. et LUONG, Q. (1996), Motion of points and lines in the uncalibrated case, *International Journal of Computer Vision* 17(1), 7-41.
- WATT, A. (1993), *3D Computer Graphics*, Addison-Wesley, Reading, MA, USA.
- WERNER, T., HERSCH, R. et HLAVAC, V. (1995), Rendering real-world objects using view interpolation, dans *Proceedings of the IEEE International Conference on Computer Vision*, Cambridge, MA, USA, pp. 957-962.
- WOLBERG, G. (1990), *Digital Image Warping*, IEEE Computer Society Press Monograph, Los Alamitos, TX, USA.

- ZHANG, Z. (1996), Determining the epipolar geometry and its uncertainty: A review, rapport technique, INRIA, Sophia Antipolis, France.
<http://www.inria.fr/robotvis/personnel/zzhang/zzhang-eng.html>.
- ZHANG, Z. (1997), A stereovision system for a planetary rover: Calibration, correlation, registration, and fusion, *Machine Vision and Applications* 10(1), 27–34.
- ZHANG, Z. (1998a), Image-based geometrically-correct photorealistic scene/object modeling (ibphm): a review, dans R. Chin et T.-C. Pong, éditeurs, *Proceedings of the Asian Conference on Computer Vision*, Vol. 2, Hong Kong, HK, pp. 340–9.
- ZHANG, Z. (1998b), On the optimization criteria used in 2-view motion analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(7), 717–729.
- ZHANG, Z. (1998c), Understanding the relationship between the optimization criteria in two-view motion analysis, dans *Proceedings of the IEEE International Conference on Computer Vision*, Bombay, India, pp. 772–777.
- ZHANG, Z., DERICHE, R., FAUGERAS, O. et LUONG, Q. (1995), A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry, *Artificial Intelligence* 78(1-2), 87–119.
- ZHANG, Z., LUONG, Q. et FAUGERAS, O. (1996), Motion of an uncalibrated stereo rig: Self-calibration and metric reconstruction, *IEEE Transactions on Robotics and Automation* 12(1), 103–113.

ZOGLAMI, I., FAUGERAS, O. et DERICHE, R. (1997), Using geometric corners to build a 2d mosaic from a set of images, dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp. 420–425.